UNIVERSIDADE ESTADUAL PAULISTA

INSTITUTO DE BIOCIÊNCIAS – CAMPUS DE BOTUCATU

Pós-graduação em Ciências Biológicas (Genética)

“Integrando citogenética e genômica em estudos comparativos entre peixes ciclídeos e outros vertebrados”.

Juliana Mazzuchelli

Tese de doutorado apresentada ao Programa de Pós-Graduação em Ciências Biológicas (Genética) do Instituto de Biociências da UNESP como parte dos requisitos para obtenção do título de Doutor.

Orientador: Cesar Martins

Botucatu-SP

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO DE AQUIS. E TRAT. DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: *ROSEMEIRE APARECIDA VICENTE*

Mazzuchelli, Juliana
Integrando citogenética e genômica em estudos comparativos entre peixes ciclídeos e outros vertebrados / Juliana Mazzuchelli. – Botucatu : [s.n.], 2012

Tese (doutorado) - Instituto de Biociências, Universidade Estadual Paulista, 2012
Orientador: Cesar Martins
Capes: 20204000

1. Genética animal. 2. Vertebrado – Evolução. 3. Citogenética.

Palavras-chave: BACs; Cichlidae; Citogenética Molecular, Cromossomos, Evolução.

"A vontade de se preparar deve ser maior do que a vontade de vencer.

Vencer será meramente a consequência de uma boa preparação."

(Autor desconhecido)

Dedico este trabalho aos meus pais, por todo apoio, amor e carinho recebido em todos os dias da minha vida!

Em primeiro lugar agradeço a Deus por tudo que tenho e tudo que sou!

Em especial aos meus pais e minha irmã Ana Paula, por depositar em mim toda confiança, pela educação dada, pela convivência familiar, e a todos da minha família pelo carinho demonstrado e pela principalmente pela compreensão nos momentos de ausência!

Ao meu marido, Willians que contribuiu de maneira essencial para me dar ânimo e vontade de querer sempre mais.

Ao meu orientador Prof. Dr. Cesar Martins, pela orientação durante esses 5 anos e meio, pelo apoio e confiança depositada durante a realização deste trabalho de doutorado e outros trabalhos desenvolvidos em parceria.

À FAPESP, pelo auxílio financeiro do projeto e pela concessão da bolsa de estudos.

Ao Prof. Dr. Thomas Kocher e todos de sua equipe na Universidade de Maryland – College Park/USA, pela biblioteca de BACs disponibilizada, por ter me recebido em seu laboratório, por todo auxílio e sugestões recebidas durante o desenvolvimento deste trabalho.

Ao Dr. Fengtang Yang e ao Instituto Sanger pela oportunidade de estágio oferecida, por todos os ensinamentos e críticas na realização da técnica de BAC-FISH que foram de fundamental importância para a realização deste trabalho e serão muito úteis no meu futuro profissional.

Aos professores da Pós-Graduação em Genética e do Departamento de Morfologia, pelo exemplo de profissionalismo, pela confiança, amizade e conselhos profissionais.

Agradeço o apoio, carinho e amizade dos amigos, técnicos e funcionários do Departamento de Morfologia em especial aos amigos do laboratório de Genômica Integrativa: Guilherme Valente, Sárah Oliveira, Bruno Fantinatti, Marcos Geraldo, Rafael Nakagima, Érika....., Marcos, Juliana Guiust, Pedro natschigal e aos ex-companheiros....

A Tatiane, Sárah e Diogo pela amizade, companherismo e sugestões em todos os momentos, tanto na realização deste trabalho, quanto na convivência do dia a dia. Pelos momentos de diversão, muitas conversas na hora do chá, que foram de fundamental importância para relaxar da rotina semanal impiedosa deste trabalho.

Às minhas companheiras de graduação da UEL, Juliana Mara, Juliana Passipieri e Leandra, que apesar da distância sempre estiveram preocupada e dispostas a ajudar e com meu trabalho.

Agradecimentos

Ao Departamento de Morfologia, ao Instituto de Biociências de Botucatu e à Universidade Estadual Paulista, pela estrutura cedida para a realização deste trabalho.

Aos funcionários da pós-graduação pela amizade, esclarecimentos em todas as solicitações.

Muito obrigada!

SUMÁRIO

RESUMO:

Os ciclídeos representam um dos exemplos mais marcantes de evolução entre os vertebrados. Por isso, muitos estudos genômicos vêm sendo desenvolvidos para este grupo, incluindo mapas genéticos de ligação, mapas físicos com base em bibliotecas de cromossomos artificiais bacterianos (BAC) e, mais recentemente, o sequenciamento completo de genomas. Bibliotecas de BACs podem ainda ser exploradas no sentido de obtenção de sondas para mapeamento citogenético, fornecendo importantes informações sobre a estrutura e evolução cromossômica das espécies e grupos. Ainda, a análise das sequências nucleotídicas dos BACs em conjunto com mapeamento *in silico* podem ser utilizadas para integrar dados físicos cromossômicos, sequências gênicas, mapas genéticos e sequências de genomas completos. Esse trabalho teve por objetivo inferir sobre a diversificação cromossômica durante a evolução dos ciclídeos utilizando BACs de bibliotecas genômicas de *Oreochromis niloticus* para o mapeamento físico citogenético e análises comparativas entre os diversos genomas disponíveis de peixes e outros vertebrados. Os resultados mostram uma forte conservação cromossômica dos marcadores/genes analisados entre as espécies de ciclídeos estudadas. Além disso, análises *in silico* de citogenética e genômica comparativa mostram grandes segmentos cromossômicos conservados não somente entre os ciclídeos, mas também entre diferentes espécies/grupos de peixes e entre espécies de vertebrados demonstrando evidências de sintenia de grandes segmentos genômicos mesmo entre grupos taxonomicamente distantes em vertebrados. Esses resultados fornecem uma excelente base para futuros estudos de caracterização citogenética e integração de mapas para estudos evolutivos em peixes, em especial na família Cichlidae.

ABSTRACT:

Cichlids represent one of the most striking examples of rapid and convergent evolutionary radiation among vertebrates. Several genomic resources have been developed for cichlid fish, including genetic linkage maps, physical mapping based on bacterial artificial chromosome (BAC) libraries and more recently and whole sequencing genomes. These BACs libraries can be explored to obtain probes for cytogenetic mapping providing important informations about chromosome structure and evolution. Additionally BAC-end sequencing analyzes in combination with *in silico* mapping can be used to integrate chromosomes, local (gene) sequences, genetic maps, and whole genome sequencing. The present study aimed to infer about the chromosomal diversification during cichlid evolution using BAC clones from *Oreochromis niloticus* libraries for cytogenetic mapping and a comparative analyzes using fish and other complete genomes available. The results show a strong conservation of number and location of hybridization signals between cichlids species. Furthermore, *in silico* comparative genomics and cytogenetic analysis showed that large chromosomal segments are conserved not only in cichlids but also among non-related fish groups and either between some vertebrates, providing evidence for large syntenic genomic segments. These results provide an excellent foundation for further cytogenetic characterization and comparative maps for chromosome evolution studies in fishes, especially in the Cichlidae family.

MATERIAIS E MÉTODOS

Capítulo I:

Capítulo II:

INTRODUÇÃO

CAPÍTULO I

CAPÍTULO II

1. INTRODUÇÃO

1.1 Integrando cromossomos e genomas: citogenômica.

Na década de 50, antes do advento de técnicas aprimoradas em biologia molecular, a citogenética forneceu as primeiras informações relativas aos genomas dos diferentes organismos (Martins et al. 2011). Naquela época, a citogenética estava baseada apenas em métodos de coloração convencional que permitiam a descrição da macro-estrutura cariotípica como o número diploide, tamanho e forma cromossômica, presença de cromossomos supranumerários ou sexuais e também polimorfismos (Graphodastky et al. 2011, Martins et al. 2011).

Por volta de 1970, técnicas de bandamento cromossômicos foram sendo implementadas (Banda C, AgNOR, Banda G, entre outros) e revelaram características mais detalhadas sobre o genoma (bandas longitudinais nos cromossomos, composição de eu/heterocromatina, regiões organizadoras de nucléolos) (Sumner 1972, 1990). Na história da citogenética o uso de fluorocromos também foi importante, sendo a quinacrina a primeira técnica moderna de bandamento utilizada (Sumner 1990). Hoje em dia os fluorocromos, principalmente DAPI e CMA3 tem sido muito usados para discriminar a composição de bases da região heterocromática e o número e localização dos sítios de DNA ribossomal (Martins et al. 2011).

Os badamentos cromossômicos foram as primeiras ferramentas usadas para a comparação de genomas, pois, espécies próximas apresentam um padrão muito similar de bandas tornando o estudo comparativo viável. Entretanto, apesar dos grandes avanços proporcionados pelas técnicas de bandamento cromossômico, elas ainda esbarravam no limite morfológico e não detalhavam o conteúdo de DNA; continuavam

defasadas frente aos grandes avanços que estavam ocorrendo na biologia molecular (Graphodatsky et al. 2011).

Após 1980, com o desenvolvimento da técnica de hibridação *in situ* fluorescente (FISH), a citogenética permitiu avanços significativos para o conhecimento do genoma com base nos cromossomos (Martins et al. 2011). A FISH possibilitou a localização de sequências nucleotídicas específicas nos cromossomos (Forster et al. 1985, Pinkel et al. 1986), definiu a transição da era clássica da citogenética à era molecular, permitindo estudos mais detalhados com a integração da informação molecular das sequências de DNA e sua localização física ao longo dos cromossomos e genomas (Schwarzacher 2003, Jiang e Gill 2006) (Figura 1). A FISH pode ser aplicada para a detecção de sequências de cópia única ou repetida, regiões específicas do cromossomo como telômeros e centrômeros, ou até mesmo cromossomos e genomas inteiros (Martins et al 2011).

Hoje em dia, estratégias avançadas vêm sendo desenvolvidas para aumentar o conhecimento sobre os genomas como um todo. A enorme quantidade de sequências de genomas completos disponíveis (http://www.genome.gov, http://genome.ucsc.edu, http://www.ncbi.nlm.nih.gov/), o desenvolvimento da bioinformática e os avanços obtidos em microscopia, facilitou a integração de dados genômicos e cromossômicos, e se tornou uma ferramenta promissora para os estudos citogenéticos, com mais perspectivas para o mapeamento físico cromossômico comparativo em alguns grupos (principalmente mamíferos) (Graphodastky et al. 2011, Martins et al. 2011) (Figura 1). Esses mapas comparativos abrem uma oportunidade de análise em muitas espécies como ferramenta para estudos evolutivos (Graphodatskii et al. 2011).

Os avanços em bioinformática permitiram a reconstrução de cariótipos baseado em análises *in silico* das sequências dos organismos modelos depositadas em alguns

bancos de dados. Esta técnica foi aplicada em estudos envolvendo humanos, galinha, zebrafish e pufferfish para a reconstrução do cariótipo ancestral de vertebrados (Kohn et al. 2006). Da mesma maneira abre-se uma nova perspectiva para análises de cromossomos e cariótipos com a "citogenética *in silico*", que permite a identificação de grupos de ligação conservados entre espécies distantes (Martins et al. 2011).

"Cytogenenomics" é um novo termo que foi usado para definir a citogenética integrando dados genômicos (Martins et al. 2011). Este novo foco da citogenética mostrou-se evidente a partir da disponibilização de bibliotecas genômicas em cromossomos artificiais bacterianos (bacterial artificial chromosome - BAC). A biblioteca de BACs é uma coleção de clones com grandes fragmentos de DNA (100-300 kb) que permite um fácil acesso ao DNA estável para manipulação (Martins et al. 2011, Tunster et al. 2011). Os BACs vêm sendo utilizados nos projetos de sequenciamento dos genomas (Lander et al. 2001, Chen et al. 2004), para estudos genômicos em geral (Klein et al. 2000, Beck et al. 2001, Gong et al. 2003), em estudos comparativos (Miyake and Amemiya 2004) e mapeamento físico cromossômico (Jiang et al. 1995, Hoskins et al. 2000, Wu et al. 2004, Romanov et al. 2005, Stiglec et al. 2007, Yasukochi et al. 2009).

A hibridização *in situ* de BACs pode permitir o reconhecimento de cada par de cromossomos individualmente, determinar as relações cromossômicas entre espécies relacionadas e também integrar mapas físicos e genéticos (Cheng et al. 2001, Pedrosa et al. 2003, Hasterok et al. 2006, Ferguson-Smith e Trifonov 2007, Griffin et al. 2007). Esses mapas integrados de genética e citogenética forneceram informações sobre a distribuição de heterocromatina, eucromatina, centrômeros e podem contribuir significativamente para a montagem das sequências genômicas (Harper et al. 2000, Cheng et al. 2001) (Figura 1). A integração de mapas genéticos tem sido

extensivamente utilizada na pesquisa, o que pode ser verificado pela quantidade de artigos publicados na última década, aplicada a diferentes espécies vegetais (Lombard e Delourme 2001, Ulloa et al. 2002, Paran et al. 2004, Yan et al. 2005, Doligez et al. 2006, Tang et al. 2009) e animais (Romanov et al. 2005, Duke et al. 2007, Kawai et al. 2007, Yasukochi et al. 2009, Di Meo et al. 2011). Em relação aos estudos em animais, apesar da maioria dos trabalhos utilizando BACs terem sido conduzidos em mamíferos, alguns trabalhos utilizaram peixes, aves, moluscos, insetos e outros grupos (Hoskins et al. 2000, Aerts et al. 2003, Romanov et al. 2005, Cnaani et al. 2007, Phillips et al. 2009, Huan et al. 2009, Ocalewicz et al. 2009, Yasukochi et al. 2009, Lee et al. 2011).

A maior dificuldade nesta técnica é a disponibilidade da biblioteca de BACs, mas com o rápido avanço nas tecnologias de sequenciamento dos genomas, a quantidade de bibliotecas disponíveis vem crescendo (Li et al. 1999, Werner et al. 1999, Quiniou et al. 2003, Thorsen et al. 2005, Katagiri et al. 2005, Stiglec et al. 2007, Watanabe et al. 2007, Wu et al. 2009) (outras bibliotecas disponíveis: http://bacpac.chori.org/home.htm). No momento, o uso de BAC no mapeamento do genoma dos peixes está em evidência, pois muitos projetos de sequenciamento estão sendo finalizados para algumas espécies de teleósteos, por exemplo, *Oreochromis niloticus*, e disponibilizados em banco de dados (www.ensembl.org), tornando viável o uso de suas bibliotecas (Kuhl et al. 2011).

O desenvolvimento de técnicas modernas de sequenciamento prometem grandes contribuições para a "citogenômica" no grupo dos peixes, integrando os dados genômicos e cromossômicos e proporcionando ferramentas para resolução de questões cromossômicas em respostas moleculares. Cromossomos microdissectados ou "flow sorted" podem ser sequenciados completamente e os dados obtidos comparados com os dados cromossômicos. Este avanço parece ser bastante eficiente para investigar uma

série de questões muito importantes na área da citogenética, como cromossomos sexuais, cromossomos B e outros polimorfismos estruturais.

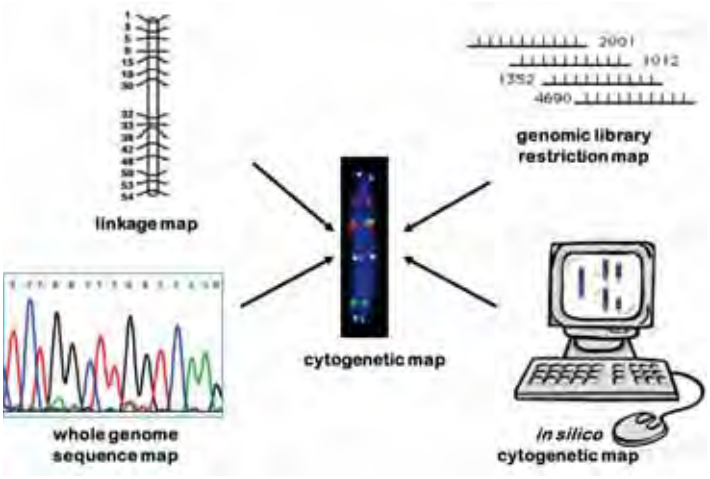


Figura 1: Integrando cromossomos, sequenciamento nucleotídico, mapas genéticos e bancos de dados para melhor compreensão do genoma.

1.2 Citogenética e genômica comparativa.

Com os recentes progressos na elucidação da sequência nucleotídica completa dos genomas das mais diversas espécies vegetais e animais (http://www.genome.gov, http://genome.ucsc.edu, http://www.ncbi.nlm.nih.gov/) tornou-se possível investigar as relações entre os genomas de diferentes espécies por análise comparativa contribuindo para a compreensão da evolução genômica como um todo.

Em detalhes, o estudo comparativo nos permite detecção de similaridades e diferenças entre genomas completos, identificação de genes ou grupos de genes envolvidos em diversas funções, identificação de genes homólogos, anotação de genes de genomas não completos, inferência de relações filogenéticas entre organismos, entre

outras abordagens (Ahn e Tanksley 1993, Keller e Feuillet 2000, Everts-van der Wind et al. 2005, Dalrymple et al. 2007, Kawai et al. 2007, Putnam et al. 2007, Reed et al. 2007, Yasukochi et al. 2009, Chapus and Edwards 2009, Sahara et al. 2007, d'Alençon et al. 2010, Di Meo et al. 2011, Nanda et al. 2011). Diante dessas análises, destaca-se o quanto o genoma eucarioto em geral se mantem conservado entre espécies do mesmo grupo, e até entre espécies de grupos distantes. A organização geral do genoma pode ter sido rearranjada por milhões de anos por meio de quebras cromossômicas, fusões, duplicações e deleções, mas ainda é possível identificar blocos menores mantendo a mesma ordem dos genes; essas regiões são denominadas regiões de sintenia (Revanna et al. 2011).

No grupo dos invertebrados, por exemplo, mapas cromossômicos identificando regiões de sintenia foram gerados para espécies de Lepidópteras (Yasukochi et al. 2007), mesmo sendo um grupo extremamente diverso com mais de 150 mil espécies identificadas. Outros estudos comparativos também evidenciaram genomas conservados e regiões de sintenia entre o genoma de *Bombyx mori* (bicho-da-seda) e as espécies de borboletas *Heliconius melpomene* e *Bicyclus anynana* (Yasukochi et al. 2006, Pringle et al. 2007, Beldade et al. 2009, d'Alençon et al. 2010, Conceição et al. 2011).

Nos peixes, muitas regiões de sintenia foram localizadas por meio das sequências terminais dos BACs entre *Oncorhynchus mykiss* (Salmonidae - truta) e as espécies de peixes modelos zebrafish, medaka e stickleback (Genet et al. 2011, Palti et al. 2011). O genoma de stickleback quando comparado com a espécie *Sparus aurata* (perciforme marinho), revela uma similaridade de aproximadamente 75% (Kuhl et al. 2011). Além disso, destacam-se as análises comparativas entre os genomas de stickleback e *Oreochomis niloticus* (Cichlidae - tilápia) (Sarropoulou et al. 2008, Soler et al. 2010), revelando uma alta similaridade, justificando o uso do genoma do

stickleback como a melhor referência para anotação do genoma de *O. niloticus* recentemente finalizado (Soler et al. 2010).

Algumas regiões de sintenia foram identificadas entre os genomas de channel catfish (*Ictalurus punctatus*) e zebrafish (Liu et al. 2009) e também com *Tetraodon nigroviridis* (pufferfish) (Wang et al. 2007). Apesar de regiões sintênicas relativamente menores quando comparadas às outras análises de sintenia obtidas para peixes descritas acima, um bom nível de conservação genômica local é detectada frente ao enorme número de rearranjos encontrados entre essas espécies.

Nas aves em geral, a evolução cromossômica é caracterizada por uma baixa taxa de rearranjo inter e intra-cromossômicos, e recentes estudos comparativos tem demostrando que muitos cromossomos tem se mantido intactos durante os 100 milhões de anos de evolução, inclusive o cromossomo sexual Z (revisado por Ellegren 2010).

Entre os mamíferos existem inúmeros trabalhos envolvendo genômica e citogenética comparativa. Em geral, apesar da grande variação existente no cariótipo deste grupo, com espécies variando o número diploide de 6 a 102, presença de cromossomos sexuais e cromossomos B (Kemkemer et al. 2009, Graphodatsky et al. 2011); muitas regiões cromossômicas conservadas tem sido identificadas abrindo possibilidades para a reconstrução do cariótipo ancestral para o grupo (Wienberg 2004, Yang et al. 2003, Murphy et al. 2005, Ma et al. 2006, Yang et al. 2006, Ferguson-Smith e Trifonov 2007, Kemkemer et al. 2009). Cromossomos totais humanos, ou segmentos cromossômicos, ou mesmo genes podem ser hibridizados nos cromossomos de outras espécies de mamíferos com alta confiança (Graphodastky et al. 2011).

Quando comparamos grupos diferentes como, mamíferos e aves, grandes regiões de sintenia são detectadas, apesar dos mamíferos apresentarem maiores taxas de rearranjos cromossômicos do que as aves (Ellegren 2010). Com os recentes progressos

obtidos no conhecimento do genoma de *Gallus gallus*, uma extensa homologia entre *G. gallus* e humanos foi identificada (Burt et al. 1999, Groenen et al. 2000, Schmid et al. 2000). Todavia, o cenário muda em relação aos cromossomos sexuais XY e ZW, os cromossomos XY humano e ZW de aves não compartilham regiões conservadas, sugerindo uma diferente origem a partir da espécie ancestral (Nanda et al. 1999, 2000, 2002).

Estudos envolvendo aves e répteis (galinha, jacaré e tartarugas), sugerem um substancial nível de divergência genômica entre esses grupos; um grande número de rearranjos cromossômicos e deleções de pequena escala são encontrados (Chapus and Edwards 2009). Além disso, o cromossomo Z de aves não apresenta sintenia com o cromossomo Z de répteis, são oriundos de autossomos diferentes do ancestral comum (Kawai et al. 2007), e revela-se conservado em pequenos blocos em diferentes autossomos (Porkoná et al. 2011).

Apesar de muito questionável a confiabilidade dos mapas comparativos entre organismos distantes evolutivamente, como mamíferos e peixes; sintenias entre humanos, peixes, insetos e cnidários foram identificadas (Putnam et al. 2007). Humanos e peixes compartilham aproximadamente 200 blocos de genes (Barbazuk et al. 2000), dentre eles, regiões sintênicas entre o cromossomo X humano e 3 cromossomos diferentes de pufferfish (Grutzner et al. 2002). Além disso, os genomas compactos das espécies de peixes pufferfish: *Fugu rubripes* (400Mb) e *T. nigroviridis* (380Mb), estão sendo usados para o mapeamento comparativo em larga escala a fim de caracterizar todo genoma humano e de outros vertebrados (Jailon et al. 2004, Mitani et al. 2006, Nakatani et al. 2007).

É óbvio de se pensar que espécies próximas, descendentes do mesmo ancestral, tenham cromossomos similares; entretanto, diante desses recentes estudos de análise

comparativa, torna-se evidente o fato de que espécies e grupos diferentes são genomicamente conservadas, mesmo depois de muito tempo separadas pela evolução (Graphodastky et al. 2011). O estudo de um maior número de espécies permitirá definir se esta conservação genômica encontrada até mesmo em espécies distantes evolutivamente é arbitrária ou se, por outro lado, a conservação de alguns segmentos reflete características funcionais da organização dos genomas. Deste modo, destaca-se a importância do uso de organismos modelos, e o aumento dos estudos que envolvam genômica comparativa.

1.3 Uma visão geral dos ciclídeos.

O grupo dos peixes é o mais antigo, numeroso e diverso dentre os vertebrados. Dentro deste grupo, a família Cichlidae (Perciformes) apresenta o maior número de espécies já relatado (Nelson 2006), sendo estimadas cerca de 3.000 espécies distribuídas principalmente pelas águas tropicais da América Central e do Sul, África, Madagascar e, algumas poucas espécies ainda são encontradas no Sudeste da Índia (Kocher 2004, Turner 2007). A família Cichlidae é um grupo monofilético composto por 4 subfamílias: Etroplinae (Ásia), Ptychochrominae (Madagascar), Cichlinae (América do Sul) e Pseudocrenilabrinae (África) (Sparks and Smith 2004, Smtih et al. 2008).

Acredita-se que a subfamília Cichlinae (ciclídeos Neotropicais) represente um grupo monofilético de origem africana (de acordo com os eventos de fragmentação de Gonduana [Pitman et al. 1993, Storey et al. 1995]) e a partir da América do Sul se espalharam pela América Central e do Norte (Murray 2001). A análise filogenética mais aceita (Smith et al. 2008) utilizou caracteres morfológicos, DNA nuclear e mitocondrial para elucidar esta questão. A análise simultânea de 6.309 caracteres, incluindo representantes de todas as grandes linhagens de ciclídeos e todos os gêneros

Neotropicais, resultou na primeira filogenia bem suportada e resolutiva para ciclídeos. A subfamília Neotropical Cichlinae é grupo irmão da subfamília Pseudocrenilabrinae (ciclídeos africanos) e foi dividida em sete tribos: Astronotini, Chaetobranchini, Cichlasomatini, Cichlini, Geophagini, Heroini e Retroculini (Smith et al. 2008).

Os ciclídeos africanos (Pseudocrenilabrinae) podem ser divididos em três grandes grupos, mas estes grupos não são reconhecidos como unidades taxonômicas válidas: (1) pelmatochromine, (2) haplochromine e (3) tilapiine (Lowe-McConnell, 1991). O grupo dos tilapiine é o grupo mais importante economicamente, cujas espécies são comumente denominadas de tilápias, e tem uma grande importância na piscicultura mundial. São espécies pertencentes aos gêneros *Sarotherodon*, *Oreochromis* e *Tilapia*, e um quarto gênero, *Danakilia*. Atualmente, a tilápia do Nilo (*O. niloticus*), é um dos peixes de água doce mais cultivado em todo o mundo, com uma produção anual de mais de 2.5 milhões de toneladas (FAO 2009 – Food and Agriculture Organization of the United Nation: http://www.fao.org/corp/statistics/en/).

Os ciclídeos em geral são muito apreciados na economia como fonte de alimento, pesca esportiva (ex: *O. niloticus* (tilápias); *Cichla* spp (tucunarés)), e aquariofilia, devido ao colorido fascinante que algumas espécies apresentam (ex: *Metriaclima*, *Etroplus*, *Symphysodon, Pterophyllum e Astronotus*, entre outras) (Axrelrod 1996, Kullander 2003). Em geral alimentam-se de uma variedade de invertebrados, peixes e plantas (Lowe-McConnell 1991, Kullander 2003). Possuem uma boca protrátil e circundada por grossos lábios, dois conjuntos de mandíbulas, uma delas forma a boca para sugar, raspar ou arrancar pequenas porções de alimento; a outra se constitui de um raro conjunto estruturado de mandíbulas na faringe poderosas e flexíveis no processamento de alimentos (Turner et al. 2007).

As mandíbulas e os dentes dos ciclídeos sofreram ação da seleção natural, sendo muito versáteis e adaptáveis para captura do alimento em nichos específicos (Albertson 2003a, 2003b). A diferenciação do aparelho de alimentação é um dos fatores que favoreceu a alta especiação dos ciclídeos nos lagos africanos (Kocher 2004).

A seleção sexual é outro fator importante neste processo de especiação para os ciclídeos africanos (Domynei 1984, Seehausen et al. 1998). O cuidado parental, assim como a guarda de ovos e larvas é marcante entre os indivíduos desta família, sendo este papel desempenhado principalmente pelas fêmeas. Os machos contribuem apenas com genes que serão herdados. Essa discrepância de investimento no cuidado parental favorece a seleção sexual realizada pelas fêmeas (Kocher 2004). Estudos revelam que a escolha do macho pela fêmea baseada na coloração apresentada pode ser observada claramente nos haplochromíneos, onde os machos possuem cores diferentes uns dos outros, adaptando-se a preferência da fêmea (Deutsch 1997, Seehausen et al. 1999, Seehausen e van Alphen 1999, Barlow 2000, Danley e Kocher 2001, Pauers et al. 2008, Pauers 2011).

Devido à diversidade de nichos ecológicos que ocupam, às estratégias de sobrevivência e às adaptações de comportamento e morfológicas que apresentam, os ciclídeos vêm sendo reconhecidos como modelos excelentes para estudos evolutivos (Lowe-McConnell 1991, Barlow 2000). A maior diversidade entre os ciclídeos é encontrada nos grandes lagos africanos, Malawi, Tanganica e Victoria, onde o número de espécies pode chegar a 1.500. Este número pode ser ainda maior, pois muitas espécies identificadas ainda não foram descritas e, além disso, muitas espécies não foram nem sequer descobertas (Turner 2007).

Os peixes ciclídeos têm claramente uma propensão a se diversificar dentro do próprio lago, e recentes estudos utilizando filogenia molecular indicaram que, dentro de

cada grande lago Africano a maioria das espécies de peixes ciclídeos evoluiu mais recentemente que a data da origem do lago e também são mais estritamente relacionadas com as outras espécies endêmicas do mesmo lago (Turner 2007).

O lago Tanganica, o mais antigo dos lagos desta região, com uma idade de 9-12 milhões de anos, possui um número estimado de 250 espécies de ciclídeos, aparenta ser oligofilético, isto é, algumas espécies de rios próximos adentraram o habitat do lago e radiaram em paralelo (Salzburger et al. 2002). O lago Malawi, é o mais diverso de todos, com 2-5 milhões de anos, possui cerca de 700 espécies; é um lago transparente e brilhante e tende a ter peixes com coloração predominantemente azul e amarela. O Lago Victoria (possui cerca de 500 ou mais espécies de ciclídeos), formado entre 250-750 mil anos atrás é mais turvo, tende a ter peixes com pigmentos vermelho e verde (Deutsch 1997, McElroy et al. 1991, Turner et al. 2001, Salzburger e Meyer, 2004 Seehausen et al. 2008, Dalton et al. 2010).

São nestes dois lagos, Victoria e Malawi, que se encontra o grupo dos haplochromíneos, considerado o grupo mais diverso dentro da família Cichlidae (Turner 2007). Os haplochromíneos da África Central e Leste, especialmente aqueles dos lagos do Rift Valley, são considerados um exemplo clássico de radiação adaptativa (Liem, 1991).

Figura 2: Diversidade do padrão dos haplochromíneos do Lago Malawi. Obtido de Turner et al. 2007.

1.4 Genética e citogenética de ciclídeos

O ciclídeos são os peixes mais estudados citogeneticamente dentro da ordem Perciformes, entretanto, os estudos atuais englobam apenas cerca de 25% do total de espécies (Feldberg et al. 2003, Oliveira et al. 2009). Os números diploides encontrados na família Cichlidae variam de 38 a 60 cromossomos. Este número está correlacionado à distribuição geográfica das espécies. Os ciclídeos africanos têm um número diploide modal igual a 44 cromossomos, já os da região Neotropical apresentam, em sua maioria, 48 cromossomos e possuem uma maior variação na fórmula cariotípica do que as espécies africanas (Revisado por Feldberg et al. 2003, Poletto et al. 2010a).

Embora não tenham sido relatados muitos polimorfismos cromossômicos neste grupo, a existência de cromossomos B foi descrita para algumas espécies como, por exemplo, as espécies sul-americanas *Gymnogeophagus balzanii* (Feldberg e Bertollo

1984), *Geophagus brasiliensis*, *Cichlasoma paranaensis*, *Crenicichla niederleinii* (Martins-Santos et al. 1995), *Crenicichla reticulata, Cichla monoculus* e *Cichla* sp. (Feldberg et al. 2004); e as espécies africanas *Astatotilapia latifasciata* do lago Victoria (citado como *Haplochromis obliquidens* em Poletto et al. 2010b), *Metriaclima lombardoi* do lago Malawi (Poletto et al. 2010a), e também foram identificados em outras 11 espécies de ciclídeos do lago Victoria (Yoshida et al. 2011).

Não ocorre a presença de cromossomos sexuais diferenciados na família e esta característica torna os estudos citogenéticos relacionados à evolução e determinação sexual interessantes, justificando o interesse por outros tipos de metodologias e marcadores ligados ao sexo (Lee et al. 2003, Lee et al. 2004, Lee et al. 2007b, Cnaani and Kocher 2008, Cnaani et al. 2008, Lee et al. 2011).

Algumas espécies africanas têm demonstrado características bem peculiares em seus cariótipos, como por exemplo, a presença de um par de cromossomos grandes bastante característicos no grupo dos tilapiíneos (revisado por Poleto et al. 2010a). Sugere-se que este cromossomo maior foi originado a partir da fusão de três diferentes cromossomos (Ferreira et al. 2010). Este par, que se destaca em tamanho dos demais cromossomos do complemento, é um excelente marcador para este grupo (Majumbar e McAndrew 1986, Poletto et al. 2010a). Apesar de apresentarem uma estrutura cariotípica extremamente conservada, muitos eventos de duplicação e inversão veem ocorrendo durante a diversificação do grupo africano Pseudocrenilabrinae. São encontradas espécies de tilapiíneos e haplochromíneos com números diploides divergentes do padrão de 44 cromossomos: *Tilapia mariae* com 2n=40 (Thompson 1981, Ferreira et al. 2010), *Tilapia sparrmanii* com 2n=42 (Thompson 1981), *Oreochromis karongae* com 2n=38 (Harvey et al. 2002), *Astatotilapia burtoni* com 2n=40 (Poletto et al. 2010a); entre outros.

Os estudos citogenéticos desenvolvidos em diversas espécies de ciclídeos baseiam-se principalmente em experimentos clássicos utilizando coloração convencional Giemsa, coloração Ag-NOR e bandamento C (Thompson et al. 1979, Thompson et al. 1981, Kornfield 1984, Feldberg e Bertollo 1985, Feldberg et al. 2003, Bettin-Pires et al. 2010; Poletto et al. 2010a); entretanto muitos estudos já foram realizados utilizando técnicas de citogenética molecular usando DNAs repetitivos como sonda.

O primeiro mapa físico gerado com sequências repetitivas nos cromossomos de *O. niloticus* foi descrito por Martins et al. (2004). Além disso, o uso de DNAs ribossomais 5S e 18S (Martins et al. 2000, Martins et al. 2002, Vicari et al. 2006, Poletto et al. 2010a), diversos DNAs repetitivos não codificantes (DNAs satélites, elementos transponíveis) (Oliveira e Wright 1998, Oliveira et al. 1999, Chew et al. 2002, Oliveira et al. 2003, Harvey et al. 2003a, Ferreira e Martins 2008, Gross et al. 2009, Mazzuchelli e Martins 2009, Teixeira et al. 2009, Ferreira et al. 2010, Gross et al. 2010, Valente et al. 2011) tem contribuído para um melhor entendimento do genoma de algumas espécies, além da sua aplicação como marcadores em estudos de evolução cromossômica. Adiciona-se estudos prévios de mapeamento de BACs contendo genes ou sequências repetitivas de DNA da tilápia do Nilo (Ferreira e Martins 2008) e avanços em mapeamento de BACs contendo genes de cópia única em espécies do gênero *Oreochromis* (Cnaani et al. 2008, Ocalewicz et al. 2009, Mota-Velasco et al. 2010).

Percebe-se que a maior parte dos avanços está direcionada a tilápia do Nilo, *O. niloticus*, devido ao seu valor econômico e a facilidade adaptativa, as tilápias se destacam como excelentes modelos experimentais para estudos em genética. Os estudos que integram dados genéticos e de citogenética molecular direcionados aos ciclídeos africanos estão no início (Cnaani et al. 2008, Mota-Velasco et al. 2010), e em relação

aos ciclídeos sul-americanos, não existe nenhum tipo de mapeamento genômico que envolva dados cromossômicos e/ou moleculares, sendo a maioria das informações genéticas existentes para este grupo, relacionadas a dados cariotípicos, localização de poucas sequências repetitivas e análises da filogenia das espécies (Farias et al. 1999, Feldberg et al. 2003, Smith et al. 2008, Gross et al. 2009, Mazzuchelli e Martins 2009, Teixeira et al. 2009, Bettin-Pires et al. 2010, Poletto et al. 2010a).

Avanços consideráveis foram obtidos em dados genômicos para a tilápia, assim como para outros peixes ciclídeos do Leste africano. Mapas genéticos têm sido publicados para *O. niloticus* (Kocher et al. 1998, Agresti et al. 2000, Lee et al. 2005), haplochromíneos do lago Malawi (Albertson et al. 2003b) e *Astatotilapia burtoni* (Sanetra et al. 2009). Estes mapas representam um importante passo para o mapeamento de locos de caracteres quantitativos que podem ser utilizados na piscicultura, em processos de melhoramento de linhagens, assim como em outros tipos de análise envolvendo a identificação de marcadores de valor na biologia básica e evolutiva dos ciclídeos. Foram desenvolvidas também bibliotecas de ESTs (Expressed Sequence Tag) para os haplochromíneos do Lago Victoria (Watanabe et al. 2004, Kobayashi et al. 2009), *A. burtoni* (Renn et al. 2004, Salzburguer et al. 2008) e para a tilápia do Nilo (Lee et al. 2010). Diversas bibliotecas em BACs tem sido construídas para a tilápia do Nilo (Katagiri et al. 2001), e também para espécies de ciclídeos dos lagos Malawi (Di Palma et al. 2007), Victoria (Watanabe et al. 2003) e Tanganica (Lang et al. 2006). Essas bibliotecas construídas em BACs representam fontes promissoras a serem utilizadas no isolamento e caracterização de genes de ciclídeos.

A perspectiva de sequenciamento completo do genoma de diferentes espécies de ciclídeos tornou-se real a partir da aprovação, no final de 2006, pela agência de fomento americana NIH (*National Institutes of Health*), da proposta de sequenciamento dos

genomas das espécies *O. niloticus*, *Astatotilapia burtoni* (Lago Tanganica), *Paralabidochromis chilotes* (Lago Victoria), e *Metriaclima zebra* (Lago Malawi) (The International Cichlid Genome Consortium 2006). Assim, o genoma completo de *O. niloticus* foi disponibilizado a comunidade no início de 2011 (http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html), e os estudos genômicos em ciclídeos têm avançado de forma espetacular com a utilização destes dados disponibilizados.

Ainda, diversos marcadores (genes e microssatélites), relacionados a sexo, foram identificados em pelo menos quatro loci distintos (grupos de ligação (LG) 1, 3, 5, e 7) (Figura 3) distribuídos em diferentes espécies de ciclídeos africanos (Lee et al. 2004, Cnaani et al. 2008, Cnaani e Kocher 2008, Ser et al. 2009). Os genes *CYP19a* e *WT1* presentes no LG1 estão envolvidos na diferenciação sexual de mamíferos e são candidatos a estarem envolvidos neste processo também em outros vertebrados (D'Cotta et al. 2001, Hossain e Saunders 2001, Chang et al. 2005). No LG3, a presença do *CLCN5*, um gene associado com o distúrbio renal, se torna interessante pela sua presença já detectada no cromossomo X de humanos (Levtchenko et al. 2007). Os genes das opsinas, presentes no LG5, são responsáveis pelo espectro de cores visualizado pelos ciclídeos e consequentemente interferem na seleção sexual e na adaptação a novos ambientes (Carleton et al. 2010)

Da mesma maneira, genes da família *SOX*, *SOX2* (LG17) e *SOX14* (LG23) exercem um papel crítico regulando a cascata de determinação sexual e desenvolvimento do sistema nervoso (Bowles et al. 2000, Kamachi et al. 2000, Kiefer 2007). Portanto, estes e outros marcadores presentes nestes grupos de ligação se mostram extremamente úteis para o mapeamento físico cromossômico em ciclídeos

visto que, pesquisas envolvendo marcadores de sexo e, determinação e diferenciação sexual neste grupo veem sendo desenvolvidas há mais de 50 anos.

1.5 Determinação sexual em Teleostei: ênfase na família Cichlidae.

Nos peixes, a determinação sexual é muito plástica e, em muitas espécies, dada por mecanismos genéticos aliados a fatores ambientais e hormonais (Baroiller e D'Cotta 2001). Podem ser encontradas espécies dioicas com escassas características sexuais secundárias, espécies com acentuado dimorfismo sexual, inversão sexual, acasalamentos complexos, cuidado parental e diversas formas de hermafroditismos (Devlin e Nagahama 2002).

Em algumas espécies podem ser encontrados diversos sistemas de cromossomos sexuais (XX/XY, ZZ/ZW, $X_1X_1X_2X_2/X_1X_2Y$, XX/XY_1Y_2, ZZ/ZW_1W_2) (Price 1984, revisado por Devlin e Nagahama 2002, revisado por Mank et al. 2006). Além disso, alguns locos autossômicos contribuem para a diferenciação sexual nas demais espécies com cromossomos sexuais indiferenciados (Kosswig 1964). Acredita-se que a determinação sexual autossômica seja a condição ancestral para o inicio da evolução dos cromossomos sexuais bem diferenciados (revisado por Mank et al. 2006). Entre as espécies que possuem cromossomos sexuais, estes parecem não ter evoluído de um par XY ancestral comum (Rexroad et al. 2008). Cromossomos XY podem ter evoluído muitas vezes de autossomos diferentes, ou uma região de determinação sexual pode ter mudado para um autossomo várias vezes. Também é possível que um novo gene de determinação sexual possa ter substituído um pré-existente em algumas espécies. Um mapeamento mais detalhado pode revelar sintenias que irão excluir ou confirmar esse cenário (Charlesworth e Mank 2010).

Os ciclídeos não apresentam cromossomos sexuais diferenciados e, estudos realizados em tilapiíneos e haplochromíneos, desvendaram que o mecanismo de determinação sexual é muito diverso e pode ser controlado por genes presentes em 4 cromossomos autossômicos diferentes que contêm grupos de ligação já conhecidos (LG 1, 3, 5, e 7) (Figura 3). Lee e colaboradores (2003) identificaram um sistema de heterogametia masculina (XY) e uma região envolvida com determinação de sexo próximo ao marcador UNH995 no LG1 da tilápia do Nilo, *O. niloticus*. Heterogametia feminina (WZ) no LG3 foram detectados na tilápia azul (*O. aureus*) (Lee et al. 2004), *O. karongae*, *T. mariae* (Cnaani et al. 2008) e *O. tanganicae* (Cnaani e Kocher 2008). Algumas famílias da tilápia azul têm sido encontradas segregando ambos os locos, e neste caso LG3 (WZ) é epistático em relação ao LG1 (XY). Análises filogenéticas não puderam determinar qual dos dois sistemas é o sistema de determinação sexual ancestral; entretanto existem evidências sugerindo que LG3 é o cromossomo sexual ancestral, e que, o controle da determinação sexual, tenha mudado para um ou mais locos presentes no LG1 (Cnaani et al. 2008).

Pelo menos dois sistemas de determinação sexual foram identificados nos ciclídeos do lago Malawi. Um sistema de heterogametia masculina (XY) no LG7 e um sistema de heterogametia feminina (WZ) no LG5 (Ser et al. 2009). Ainda podem ser encontradas espécies, como *Metriaclima pyrsonotus*, que possuem dois locos envolvidos na determinação sexual; neste caso o loco WZ é epistático sobre o XY; e também espécies que não tiveram seu mecanismo de determinação relacionado com nenhum destes locos (Ser et al. 2009).

Figura 3: Locos de determinação sexual identificados em quatro grupos de ligação (LG1, LG3, LG5, LG7) para diversas espécies de ciclídeos africanos. Os marcadores destacados em vermelho possuem BACs identificados para estas regiões.

Os ciclídeos representam um dos exemplos mais marcantes de evolução entre os vertebrados (Turner 2007) se tornando alvo de um grande número de pesquisas científicas em especial, estudos genéticos. Com o genoma de *O. niloticus* disponível, abre-se novas oportunidades para a integração de toda informação disponível para esta família (dados citogenéticos, morfólogicos e genômicos) proporcionando uma nova perspectiva nos estudos cromossômicos de peixes em geral, permitindo uma análise mais apurada do genoma e da sua dinâmica evolutiva.

2. OBJETIVOS

2.1 Objetivo geral

O presente trabalho teve como objetivo contribuir com a elucidação dos processos envolvidos na diversificação cromossômica/genômica de peixes ciclídeos utilizando como marcadores citogenéticos BACs dos grupos de ligação 1, 3, 5, 7, 17 e 23, integrando os dados obtidos com dados genômicos.

2.1 Objetivos específicos

1- Comparar e entender a diversificação da estrutura cariotípica entre espécies de ciclídeos a partir da construção de mapas cromossômicos comparativos utilizando BACs;

2- Entender os mecanismos envolvidos na diversificação cromossômica/genômica dos ciclídeos integrando os dados genômicos presente em bancos de dados a dados citogenéticos obtidos para os ciclídeos;

3- Entender os processos de diversificação genômica entre ciclídeos e outros vertebrados utilizando análise de sintenia.

3. MATERIAIS E MÉTODOS

3.1 Espécies utilizadas

Foram obtidas preparações cromossômicas de espécies de ciclídeos africanos, asiático e sul-americanos, procedentes de diversas fontes (Tabelas 1 e 2) (Licença do IBAMA para coleta número 15729-1). Os exemplares foram adequadamente estocados nas coleções do Laboratório Genômica Integrativa da UNESP e na Tropical Aquaculture Facility - Universidade de Maryland/USA.

Tabela 1: Ciclídeos africanos (Pseudocrenilabrinae) e asiático (Etroplinae) analisados.

Subfamílias e espécies	Origem	Grupo	N
Etroplinae			
Etroplus maculatus	Loja de aquariofilia, Botucatu, SP		06
Pseudocrenilabrinae			
Oreochromis aureus	Tropical Aquaculture Facility, UMD, USA		04
O. mossambicus	Tropical Aquaculture Facility, UMD, USA		04
O. niloticus	Rio Tietê, Botucatu, SP, Brazil Instituto de Aquacultura, Stirling University, Stirling, Scotland. Tropical Aquaculture Facility, UMD, USA	**Tilapiines**	25
O. tanganicae	Tropical Aquaculture Facility, UMD, USA		01
Tilapia mariae	Tropical Aquaculture Facility, UMD, USA		06
T. mamfe	Tropical Aquaculture Facility, UMD, USA		01
Hemichromis bimaculatus	Loja de aquariofilia, Botucatu, SP, Brazil	**Hemichromines**	06
Astatotilapia burtoni	Tropical Aquaculture Facility, UMD, USA		03
A. baenschi	Tropical Aquaculture facility, UMD, USA		03
Cynotilapia afra	Tropical Aquaculture facility, UMD, USA		01
Gephyrochromis moorii	Loja de aquariofilia, Botucatu, SP	**Haplochromines**	03
Haplochromis livingstonii	Loja de aquariofilia, Botucatu, SP		01
A. latifasciata	Loja de aquariofilia, Botucatu, SP		80
Labeotropheus trewavase	Tropical Aquaculture Facility, UMD, USA Loja de aquariofilia, Botucatu, SP		09

Melanochromis auratus	Tropical Aquaculture Facility, UMD, USA Loja de aquariofilia, Botucatu, SP	02
Metriaclima barlowi	Tropical Aquaculture facility, UMD, USA	06
M. gold *zebra*	Tropical Aquaculture facility, UMD, USA	04
M. lombardoi	Tropical Aquaculture facility, UMD, USA Loja de aquariofilia, Botucatu, SP	22
M. pyrsonotus	Tropical Aquaculture facility, UMD, USA	08
Pseudotropheus tropheops	Loja de aquariofilia, Botucatu, SP	01
P. zebra	Loja de aquariofilia, Botucatu, SP	01
P. sp	Loja de aquariofilia, Botucatu, SP	01

Tabela 2: Ciclídeos sul-americanos (Cichlinae) analisados

Tribo	Espécies	Origem	N
Cichlini	*Cichla temensis*	Rio Tocantins, Tucuruí, TO	17
	C. orinocensis	Rio Orinoco, Caicara, Venezuela	01
	C. piquiti	Rio Araguaia, São Felix do Araguaia, MT	04
	C. kelberi	Rio Araguaia, São Félix do Araguaia, MT Rio Tietê, Bariri, SP	12
Retroculini	*Retroculus lapidifer*	Rio Araguaia, Barra do Garças, MT	02
Astronotini	*Astronotus ocellatus*	Rio Tietê, Barra Bonita, SP Lago Catalão, Manaus, AM	11
Chaetobranchini	*Chaetobranchus flavescens*	Rio Araguaia, São Félix do Araguaia, MT	01
Geophagini	*Apistogramma borellii*	Logoa Comprida, Aquidauana, MS	05
	Biotodoma cupido	Rio Araguaia, Barra do Garças, MT Rio Araguaia, São Félix Araguaia, MT	07
	Crenicichla lepidota	Lagoa Comprida, Aquidauana, MS	05
	C. strigata	Rio Araguaia, Barra do Garças and São Félix do Araguaia, MT	03
	C. britskii	Córrego Olaria, Poloni, SP	01
	C. aff *britskii*	Córrego Olaria, Poloni, SP	01
	C. aff *haroldoi*	Córrego Olaria, Poloni, SP	01
	Geophagus brasiliensis	Córrego Olaria, Poloni, SP Córrego Araqua, Botucatu, SP Rio Bonito, Barra Bonita, SP Rio Paraitinguinha, Salesópolis, SP	07

	G. proximus	Rio Araguaia, Barra do Garças,MT	04
	G. cf *proximus*	Rio Tietê, Buritama, SP Rio Engenheiro Taveira, Araçatuba, SP	04
	G. surinamensis	Rio Orinoco, Caicara, Venezuela	03
	Satanoperca jurupari	Rio Araguaia, Barra do Garças, MT Rio Araguaia, São Félix do Araguaia, MT	16
Cichlasomatini	*Aequidens plagiozonatus*	Lagoa Comprida, Aquidauana, MS	09
	A. tetramerus	Rio Araguaia, Barra do Garças, MT Rio Araguaia, São Félix do Araguaia, MT	09
	Cichlasoma facetum	Córrego Campo Novo, Bauru, SP Rio Paraitinguinha, Salesópolis, SP	06
	C. nigrofasciatum	Loja de aquariofilia, Botucatu, SP	13
	C. paranaense	Córrego Carrapato, Penápolis, SP Córrego Batata, Miracatú, SP Córrego Faú, Miracatú, SP	08
	Laetacara dorsigera	Rio Bahia, Pracinha, PR	01
Heroini	*Heros efasciatus*	Rio Araguaia, Barra do Garças, MT Rio Araguaia, São Félix Araguaia, MT	03
	Mesonauta festivus	Rio Araguaia, Barra do Garças, MT Rio Araguaia, São Félix Araguaia, MT	10
	Parachromis managuensis	Loja de aquariofilia, Botucatu, SP	01
	Pterophyllum scalare	Loja de aquariofilia, Botucatu, SP	04
	Symphysodon aequifasciatus	Loja de aquariofilia, Botucatu, SP	02

3.2 Obtenção dos cromossomos mitóticos.

Preparações metafásicas foram obtidas seguindo a metodologia adaptada para peixes por Bertollo et al. (1978), com modificações de acordo com as descrições que seguem: Injetar intraperitonealmente colchicina 0,025% na proporção de 1 ml para cada 100 g de peso do animal. Deixar o peixe em aquário bem aerado por 40 minutos. Em seguida sacrificar o peixe e retirar a porção anterior do rim transferindo-a para uma solução hipotônica de KCl 0,075 M (6-8 ml). Divulsionar bem o tecido com o auxílio de uma seringa de vidro. Retirar o sobrenadante (suspensão celular) com o auxílio de uma pipeta Pasteur e colocar em tubo de centrífuga. Incubar a suspensão celular obtida em estufa a 37 °C por 23 minutos. Pré-fixar com 6 gotas de metanol:ácido acético (3:1) e ressuspender o material pipetando bem devagar por 100 vezes. Deixar descansar por 5 minutos, adicionar fixador até encher o tubo e ressuspender. Centrifugar por 10 minutos a 800 rpm. Desprezar o sobrenadante e completar para 6 ml com fixador pipetando por mais 100 vezes. Centrifugar por 10 minutos a 1000 rpm, desprezar o sobrenadante e completar novamente para 6 ml de fixador, repetindo essa lavagem por mais duas vezes. Após a última lavagem, diluir o material acrescentando fixador, de forma que este apresente um aspecto um pouco turvo. Preparar as lâminas que deverão estar previamente aquecidas em banho-maria a 60 ^{o}C.

3.3 Busca por BACs de interesse.

Os esboços dos grupos de ligação relacionados com a determinação sexual nos ciclídeos, LG1, 3, 5 e 7 (Figura 3) e também os LGs contendo genes da família SOX, LG17 e 23, foram obtidos por meio do Cmap Viewer (http://dev.cichlid.umd.edu/cmap/). Os BACs de interesse para hibridação *in situ* foram selecionados da biblioteca de *O. niloticus* de acordo com sua posição no mapa de

ligação (Tabela 3, Figura 3). Para se obter o número de identificação do BAC (BAC ID), foi realizada uma análise comparativa entre o genoma de *O. niloticus* e stickleback no Cichlid genome browser (http://cichlid.umd.edu/cgi-bin/gb2/gbrowse/stickleback/). Por meio deste número foi possível localizar o clone correspondente ao BAC na biblioteca estoque de *O. niloticus* e fazer sua purificação.

Tabela 3: BAC-clones de *O. niloticus* purificados e utilizados como sondas no mapeamento cromossômico.

Marcador	LG	BAC ID	gene	Observações
UNH995	LG1	b04TI071H11		Cnaani et al. 2008, RS
CYP19A1	LG1	b04TI008J05	Ovarium aromatase	
	LG1	b04TI002B08	Ovarium aromatase	
WT1	LG1	b03TI091I08	WT	
GM354	LG3	b03TI066P02		Cnaani et al. 2008, RS
GM204	LG3	b04TI071O04		Cnaani et al. 2008, RS
UNH180	LG3	b04TI056G07		Cnaani et al. 2008, RS
UNH115	LG3	b03TI086K09		Cnaani et al. 2008, RS
CLCN5	LG3	b04TI076F11		Cnaani et al. 2008, RS
Trp1	LG3	b03TI073M01		
UNH106	LG3	b03TI088C14		
GM526	LG3	b03TI067N14		
	LG5	b04TI053F24	ATPase	3.65 Mb Tetr 11
	LG5	b04TI010O22	Green Opsin	RS
	LG5	b04TI075I09	Blue/red Opsin	
	LG5	b04TI006L21	c-ski	
	LG7	b03TI050E01	RERG	
	LG7	b03TI079D23	KCNE1L	RS
	LG7	b03TI080A15	DUSP6	
	LG7	b03TI081O07	IGFBP2	
UNH179	LG7	b04TI036P14		
UNH896	LG7	b04TI035B08		
SOX14	LG17	b03TI079I04	SOX14	Cnaani et al. 2007
SOX2	LG23	b04TI053B06	SOX2	Cnaani et al. 2007

RS: Presença de sequências repetitivas.

3.4 Obtenção de BACs para utilização como sondas cromossômicas.

Extração e purificação:

A extração dos BACs foi realizada com o Kit PhasePrep™ BAC DNA – Sigma (Product Code NA 0100) segundo orientações do fabricante para o procedimento em micro escala (Micro Scale preparation): Inoculou-se 50µl dos BACs em 50ml de meio LB (peptona 1%/NaCl 0,17M, extrato de levedura 0,5%, pH 7,5) líquido, contendo clorafenicol (1µl (20mg/ml) para 1ml de meio líquido) deixando crescer durante 14 hs a 37 ^{o}C em agitação de 250 rpm. Após este período, seguiu-se as recomendações do protocolo Kit PhasePrep™ BAC DNA para extração, remoção de resíduos e precipitação do DNA.

Para verificar a concentração e a qualidade dos clones purificados, as amostras foram analisadas e quantificadas em gel de agarose 1% e no espectrofotômetro NanoVue – GE Healthcare Life Science.

Amplificação do DNA genômico do BAC:

O próximo passo consistiu na amplificação do DNA do BAC extraído na etapa anterior utilizando o kit GenomePlex® Complete Whole Genome Amplification (WGA) (Sigma) segundo recomendações do fabricante.

Preparo da sonda - Marcação do DNA:

Os produtos de amplificação obtidos na etapa acima foram marcados através do kit de reamplificação GenomePlex® WGA Reamplification (Sigma) com algumas modificações: Para a marcação da sonda foi necessário fazer um mix de dNTP contendo 70% de um dos nucleotídeos (ex: 70% dTTP*) e adicionou-se o nucleotídeo marcado com moléculas específicas ou mesmo diretamente com fluorocromos (Biotina, Digoxigenina, Cy3, Cy5, FITC ou Texas Red, Orange ou Green-dUTP).

*dNTP mix para marcação - Mix com 70% de dTTP: dATP, dCTP, dGTP (2 mM cada), 1.3 mM dTTP.

Reação para marcação da sonda:

- 94ºC – 3 minutos
- 17 ciclos: 94 ºC – 15 segundos

  65 ºC – 5 minutos
- 4 ºC

Os produtos da reação de amplificação anterior foram fragmentados pela adição de 5µl de DNAseI (1µg/µl) e incubados por 1 hora e 30 minutos a 15ºC. Em seguida, o tamanho dos fragmentos gerados foi visualizado em gel: uma condição ideal para os fragmentos é que estejam entre 100 e 500 pares de bases (pb). A reação foi interrompida pela adição de 1µl de EDTA 0,5M e aquecimento por 10 minutos a 65ºC para inativar a enzima e prevenir a degradação da sonda.

3.5 Hibridação *in situ* fluorescente (FISH): Protocolo para BACs de acordo com Fengtang Yang (Sanger Institute, Inglaterra) (comunicação pessoal com modificações)

Tratamento pré-hibridação

1. Incubar as lâminas por 3 min em solução de pepsina 0.01% a temperatura ambiente;
2. Lavar em 2x SSC durante 5 min a temperatura ambiente;
3. Repetir a operação;
4. Desidratar as lâminas em série alcoólicas (70, 85 e 100 %) por 3 min cada;
5. Secar as lâminas em um bloco quente ou estufa a 65ºC por 1 hora.

Pré-hibridação

6. Desnaturar o DNA cromossômico com formamida 70%/2 X SSC a 65ºC por mais ou menos 1 min (depende da espécie utilizada e qualidade do material);
7. Desidratar o material em série alcoólica 70%, 85% e 100% durante 3 min cada. Obs. O álcool 70% deverá estar a –20ºC;
8. Secar bem.

Hibridação

9. Preparar a câmara úmida a 37ºC;
10. Montar cada lâmina com a solução de hibridação, cobrir com lamínula, selar e incubar overnight a 37ºC;

    Tampão de hibridação (62.5% de Formamida):

    - 25 µl Formamida 100%;
    - 10 µl Sulfato de Dextrano 50%;
    - 5 µl de 20xSSC;
    - Volume final 40 µl.

    Mix de hibridação: Usar 10 µl do tampão e mais 1 µl de cada sonda se utilizada na mesma espécie; 1,5 µl de sonda se utilizada em espécies diferentes (hibridação cruzada).

    Desnaturação da sonda a 95ºC por 10 min.

Lavagens

11. Lavar 2 vezes em formamida 50%/2xSSC pH 7.0 a 42 ºC durante 5 min cada;
12. Lavar as lâminas 2 vezes em 2xSSC a 42 ºC, por 5 min cada.

Detecção e amplificação do sinal

13. Incubar as lâminas por 20 minutos a 37 ºC em câmara escura - tampão 4xSSC / Tween 20; 1:400 (Fitc 1mg/ml); 1:200(anti-dig rodhamine 200ng/ul);
14. Lavar 2 x 5min com 4xSSC/Tween 20; 42 ºC;

15. Incubar com anti-avidina (2 µl anti-avidina/40 µl de 4xSSC/Tween 20) durante 20 min em câmara úmida e escura, 37ºC;
16. Lavar 2 x 5 min com 4xSSC/Tween 20; 42 ºC;
17. Incubar as lâminas por 20 minutos a 37ºC na câmara escura - tampão 4xSSC/Tween 20; 1:400 (Fitc 1mg/ml);
18. Lavar 2 x 5min com 4xSSC/Tween 20.

Montagem da lâmina com DAPI

19. Misturar 20 µl de antifading mais 1 µl de dapi (0,2 mg/ml);
20. Colocar 20 µl da mistura por lâmina e cobrir com lamínula. Guardar no escuro.

Processamento das imagens

Os cromossomos metafásicos foram analisados em um fotomicroscópio de fluorescência Olympus BX 61. As imagens foram capturadas com uma câmera digital (Olympus DP71) e com o programa Image-Pro MC 6.0 e processadas por meio do programa Adobe Photoshop 7.0. Algumas imagens digitais foram adquiridas com o sistema CytoVision (Applied Imaging Corp.) com uma câmera CCD acoplada a um microscópio Zeiss durante o estágio realizado no Instituto Sanger.

3.6 Análise de genômica comparativa

As sequências dos marcadores dos LG1, 3, 5 e 7 e dos genes *SOX* presentes nos LG 17 e 23, foram obtidas por meio do Genbank - National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/nuccore) e localizadas no genoma de *O. niloticus* por meio do banco de dados: http://cichlid.umd.edu/blast/blast.html (RRS5KB-p10-noN.e60.c0.p60-contigs.fa).

Utilizando a referência individual de cada BAC (BAC ID) foi possível realizar análises comparativas entre o genoma de *O. niloticus* e o genoma das espécies de peixes *Oryzias latipes* (medaka), *Tetraodon nigroviridis* (pufferfish) e *Gasterosteus aculeatus* (stickleback) presentes na versão mais recente do banco de dados BouillaBase (www.BouillaBase.org) no link "Genome browsers" - Comparação de genomas (http://cichlid.umd.edu/cichlidlabs/kocherlab/genomebrowsers.html). Por meio deste banco de dados é possível estabelecer uma comparação entre esses genomas utilizando as sequências terminais dos BACs de *O. niloticus* previamente sequenciadas.

As análises de similaridade entre os demais vertebrados foram feitas utilizando Megablast no Basic Local Alignment Search Tool (BLAST: NCBI - http://blast.ncbi.nlm.nih.gov/Blast.cgi) e a localização cromossômica dos genes/marcadores utilizados, assim como as análises de regiões de sintenias entre peixes e as demais espécies, foram determinadas utilizando informações disponíveis nos bancos de dados: NCBI Map Viewer (http://www.ncbi.nlm.nih.gov/genomes); Instituto Sanger Ensembl Database (http://www.ensembl.org), BouillaBase (www.BouillaBase.org) e Genomicus (http://www.dyogen.ens.fr/genomicus/). Cada sequência/gene foi incluída individualmente na busca e a localização cromossômica foi dada para as espécies que já tem o genoma depositado e anotado.

4.1 Capítulo I:

Integrating cytogenetics and genomics in comparative evolutionary studies of cichlid fish

Juliana Mazzuchelli[1], Fengtang Yang[2], Thomas D Kocher[3], Cesar Martins[1,*]

[1] Department of Morphology, Bioscience Institute, UNESP - São Paulo State University, 18618-000, Botucatu, SP, Brazil.

[2] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

[3] Department of Biology, University of Maryland, College Park, MD, 20742, USA.

Key words: Cichlidae, genome evolution, molecular cytogenetics, chromosome, linkage groups, BACs

Manuscrito a ser submetido para a revista BMC Genomics.

**Corresponding author*

JM: mazzuchelli@ibb.unesp.br

FY: fy1@sanger.ac.uk

TDK: tdk@umd@edu

CM: cmartins@ibb.unesp.br

Abstract

Background: The availability of hundreds of eukaryotic genomes recently sequenced opens new avenues for the integration of cytogenetics and genomics in comparative and evolutionary studies. Cytogenetic mapping offers means to define regions in the genome that are not well clarified even after strong nucleotide sequence efforts and also to identify conserved synteny shared between genomes. An efficient way to perform comparative cytogenetic mapping is based on the application of BAC clones as probes for chromosome *in situ* hybridization. In this report, to address knowledge about the genome evolution in cichlid fish, BAC clones from *Oreochromis niloticus* library covering the LG 1, 3, 5 and 7 were used to map the chromosomes of several African cichlids. The cytogenetic mapping data were also integrated to nucleotide sequence information of BAC-end sequences and comparatively analyzed against the genome of the fish species, *Tetraodon nigroviridis* (pufferfish), *Gasterosteus aculeatus* (stickleback) and *Oryzias latipes* (medaka). Also, using currently available public genomic databases, comparative analyses were also extended to other vertebrates.

Results: A strong conservation of the location of BACs of LG1, 3, 5 and 7 was observed among cichlid species, preserving the synteny of the markers of each LG. Comparative analysis also identified large genomic blocks conserved to several fish groups non-related to cichlids and also in several other non-fish vertebrates.

Conclusions: Although fishes are known to have plastic genomes with high rates of chromosomal rearrangements and probably low rates of synteny, our results evidence that large syntenic chromosome segments have been maintained conserved during evolution. Additionally, our current cytogenetic mapping efforts integrated to genomic approaches conduct to a new perspective for the chromosome biology area, providing the tools to investigate questions of the chromosomal scale into molecular answers**.**

Introduction

BAC genomic libraries have demonstrated useful for many aspects of molecular and genomic studies, such as the positional cloning of genes (Stiglec et al. 2007a), comparative studies of synteny and gene organization among different species (Goldammer et al. 2009), as well as for local or whole genome physical and genetic mapping and sequencing (Katagiri et al. 2005). The potential of using BACs for studying animal genomes has increased, because several good quality genomes are already available in public databases (see the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov/, for instance).

In cytogenetic research and chromosome mapping, BAC clones have also been used successfully as probes for FISH and the application of this approach has increased significantly in the last years in the animal cytogenetic field bringing up the possibility to refine the chromosome analysis under the focus of advanced molecular cytogenetics (Sahara et al. 2007; Freeman et al. 2007; Cnaani et al. 2008; Phillips et al. 2009; Ellegren 2010; Mota-Velasco et al. 2010; Di Meo et al. 2011; Mazzuchelli et al. 2011). The potential of BAC-FISH for studying fish genomes has increased after considerable progress has been made in developing genomic resources for several species, including members of the Cichlidae family.

Cichlid fish represent a monophyletic group classified in 4 subfamilies: Etroplinae (Indian and Madagascar), Ptychochrominae (Malagasy), Cichlinae (Neotropical region) and Pseudocrenilabrinae (African) (Sparks and Smith 2004, Smith et al. 2008). The karyotypes of more than 135 species of cichlids have been determined, and the diploid number ranges from 2n = 32 to 2n = 60. African cichlids have a modal diploid number of 44 chromosomes whereas the Neotropical cichlids 2n = 48 chromosomes (reviewed by Poletto et al. 2010). African Pseudocrenilabrinae cichlids

can be separated in three major groups (hemichromines, haplochromines and tilapiines), but these groups are not recognized as valid taxonomic units (Lowe-McConnell, 1991). The African groups of lakes Victoria, Malawi and Tanganica represent a classical example of extensive and rapid radiation of high interest for the evolutionary biology (Kocher 2004).

Genetic maps have been published for the Nile tilapia *Oreochromis niloticus* (tilapiine) (Lee et al. 2005), Lake Malawi haplochromines (Albertson et al. 2003), and *Astatotilapia burtoni* (Sanetra et al. 2009). There are also extensive collections of ESTs for Lake Victoria haplochromines (Watanabe et al. 2004, Kobayashi et al. 2009), *A. burtoni* (Renn et al. 2004, Salzburguer et al. 2008) and *O. niloticus* (Lee et al. 2010). Several BAC libraries have been constructed for Nile tilapia (Katagiri et al. 2001), and also for haplochromine cichlids from lakes Malawi (Di Palma et al. 2007), Victoria (Watanabe et al. 2003) and Tanganica (Lang et al. 2006). Altogether, these genomic resources have allowed advances concerning the investigation of several aspects of cichlid's biology, including sex determination (Ezaz et al. 2004; Lee et al. 2005; Lee e Kocher 2007; Lee et al. 2011). Tilapiines have an XY system on LG1 or a ZW system on LG3 (Cnaani et al. 2008, Cnaani and Kocher 2008). However, at least two distinct genetic sex determination systems in the Lake Malawi cichlids are found: a XY sex-determination system on LG7 and a ZW system on LG5 (Ser et al. 2009).

Additionally, the genome of *O. niloticus* has been recently sequenced (see Cichlid Genome Consortium at http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html) opening the opportunity to integrate nucleotide sequence information and other genetic data. In this context, we explore in the present work comparative analysis of cichlid fish based in the integration of genomic and molecular cytogenetic. Furthermore, the integrated genomic/cytogenetic

information obtained for cichlids was also comparatively analyzed with other vertebrates. Our results provide evidences for extensive synteny of chromosomal segments among Pseudocrenilabrinae cichlids and either between cichlids and other vertebrates. Such information is promising in the establishment of a framework for additional genome-wide studies.

Materials and methods

Animals and sampling

Cichlids from Lake Malawi were collected from the wild from 2005-2008 and maintained in the Tropical Aquaculture Facility (TAF) of the University of Maryland (UMD), College Park, MD, USA. Additional African and Asian species of uncertain origin were obtained from commercial sources in Botucatu, SP, Brazil, and South American species were collected from the wild in several Brazilian rivers (Table 1) and maintained in the fish room of the Laboratório Genômica Integrativa (FR-LGI) at São Paulo State University (UNESP), Botucatu, Brazil. All the specimens examined were fixed in formaldehyde and then stored in alcohol in the fish collections of TAF-UMD and FR-LGI.

Bacterial artificial chromosome (BAC) clones and probe labeling

BAC clones containing specific markers of LG1, 3, 5 and 7 (Table 2) were obtained from a BAC library of the Nile tilapia, *O. niloticus*, which was developed previously (Katagiri et al. 2001), and were used as probes for FISH. BAC extraction was conducted using the PhasePrep®^{TM} BAC DNA Kit (Sigma-Aldrich, St Louis, MO, USA) according to supplier's protocol. The BAC clones were labeled with biotin,

digoxigenin coupled nucleotides (Roche Applied Sciences, Indianapolis, IN, USA), CY3- and CY5-avidin (GE-Healthcare, UK) using whole genome amplification (WGA2 & 3 kits) (Sigma-Aldrich), according to the supplier's protocol. After that, a DNAse-I concentration was titrated to yield labeled DNA fragments ranging from 100 to 500 bp products. For multicolor FISH was used 16 µl of probe mixture containing: 10 µl of hybridization mixture (62,5% deionized formamide, 12,5% of 20XSSC, 25% of dextran sulfate 50%), 4 µg of blocking solution (salmon/herring sperm DNA) and 100 ng of each probe was prepared. The solution was denatured for 10 min at 65°C and immediately cooled on ice.

Chromosome preparation and FISH procedure

Mitotic chromosomes of cichlid species belonging to the Pseudocrenilabrinae (including representatives of tilapiine, haplochromine and hemichromine groups), Cichlinae and Etroplinae subfamilies (Table 1) were prepared from anterior kidney cells with in vivo colchicine treatment (Bertollo et al. 1978). The slides with chromosomes were air-dried, treated with pepsin (0.01% in 10 mM HCl) and dehydrated in an ethanol series one day before use. The slides were denatured in 70% formamide/2xSSC, pH 7 for 40s, and dehydrated in an ice-cold ethanol series. The probe mixture was hybridized under a 24 x 50 mm cover slip in a 37°C moist chamber for 48 h. Slides were washed two times for 5 min in 50% formamide/2xSSC, pH 7 at 43°C under agitation, then 10 min in 2xSSC, pH 7 at 42°C under continuous agitation. For undirected labeled probes, the hybridization signals were detected with avidin-FITC and rhodamine-anti-DIG (Roche Applied Sciences, Indianapolis, IN, USA), according to the supplier's protocol. After three washes of 2 min in phosphate buffer detergent (4xSSC/1% Tween-20), slides were mounted with antifade solution containing DAPI (Vectashield mounting

medium). Results were recorded with an Olympus BX61 microscope equipped with an Olympus digital camera DP71 and the software Image-Pro MC 6.0.

Comparative genomic database analyses

The comparative analyses between *O. niloticus* BAC-end sequences and *Oryzias latipes* (medaka), *Tetraodon nigroviridis* (pufferfish) and *Gasterosteus aculeatus* (stickleback) were done based on the BAC ID (Table 2) of the clones used for FISH against the updated version of BouillaBase database (www.bouillabase.org) at Comparative Genome Browsers (http://cichlid.umd.edu/cichlidlabs/kocherlab/genomebrowsers.html). The BAC ID was the start point to detect a landmark or region of similarity in the databank source. The identification of genomic positions of genes located in LG1, 3, 5 e 7 of *O. niloticus* in several other vertebrates were determined using the currently available public genomic databases Sanger Institute Ensembl Database (http://www.ensembl.org) and Genomicus genome browser (http://www.dyogen.ens.fr/genomicus/). Once a query gene name has been entered in the search box, the position of the gene in all species that have their genome available and annotated is retrieved. The accession numbers of all genes retrieved and used in the comparative analyze is available at Additional file 1

Results

BAC-FISH mapping

The diploid chromosome number and morphology were analyzed under Giemsa-stained metaphases (Table 1) and confirm the previous chromosome data available for cichlids (revised in Feldberg et al. 2003 and Poletto et al. 2010). Metaphases of males

and females of *O. niloticus* (the species source of the BAC clones used for FISH) were used to start the BAC-FISH mapping experiments (Figure 1). BACs for the LG1 mapped on the long arm of a small st/a chromosome pair (Figure 1A). The largest chromosome of *O. niloticus* evidenced labeling of BACs of LG3 (Figure 1B). The BAC markers of LG5 mapped on a medium st/a chromosome different from that containing the LG1 (Figure 1C), and BAC markers of LG7 maps the second largest pair (Figures 1D). No differences were found between males and females for all BAC-probes hybridized.

A comparative mapping including eight other Pseudocrenilabrinae species was carried out to elucidate the evolutionary history of chromosomes carrying LG1, 3, 5 and 7. At least two individuals of each species was used in BAC-FISH experiments and the signals pattern observed for LG1, 3, 5 and 7 appears conserved in number and position when compared with the reference species *O. niloticus* (Figure 2; See Additional file 2). In South American cichlids (Cichlinae), none of the BAC probes produced identifiable chromosomal signals. Also no signal was observed in the Asian Etroplinae species, *Etroplus maculatus*.

Besides the conservation of studied LGs, differences were observed in the morphology of the chromosomes among the species (Figure 2): *Astatotilapia burtoni* shows a different pattern for LG5 which is located on the short arm of a m/sm chromosome instead of a smal st/a chromosome as observed in the other cichlids. The only BAC signal observed in the short arms of the chromosomes were related to repeated sequences as observed for the UNH115 marker in LG3 in *Metriaclima lombardoi* (Figure 2; Additional file 2). Additional small variation in the chromosomal position of the markers was also observed (See Additional file 2), but they could be a consequence of polymorphic patterns among the analyzed individuals.

Blocking hybridization conditions were used for all BACs that mapped to euchromatin and for BACs that had a lot of repeated DNAs (Table 2). Additionally, the LG3 mapping confirmed the presence of a lot of repetitive DNA in the end of the largest chromosome of all species analyzed (Figure 2).

The selected BACs were checked if the FISH mapping results were in agreement to the anchoring marker positions on LGs. Each group of BACs from a particular linkage group was checked to hybridize to the same chromosome using at least two markers from each LG. The chromosomal positions of all selected BACs were correlated with their corresponding order on their linkage maps; this enabled the association of LGs with chromosomes and to orientate their position according to the short/long arm of the corresponding chromosomes. Using the convention that the short arm is oriented north and the long arm south, the north/south orientation of the current *Oreochromis* LG1 match with the short/long arm (Figure 3); however, the orientation of linkage groups 3, 5 and 7 should be inverted in relation to the order of molecular markers previously published (Figures 4, 5 and 6).

In silico comparative analyses

Comparative genomic analyzes of LG1 at Bouillabase genome browser suggest conserved pattern of LG1 markers in medaka chromosome 3, stickleback chromosome 2, and pufferfish chromosome 5 (Figure 3) (Table 3). Additionally using Ensembl and Genomicus database, the chromosomal position of the genes *WT1b* and *CYP19A1*, mapped in LG1, was comparative analyzed among fish and also other vertebrates (Figure 3). The present analyses show that both genes are in the chromosome 18 of *Danio rerio* (zebrafish) (Figure 3) separated by 6Mb. In other vertebrates the results are

very divergent and these genes appear in different chromosomes as observed in chicken and primates, where both markers are in chromosomes 10 and 5, and chromosomes 15 and 11, respectively (Figure 3).

LG3 appeared conserved with at least two correspondent markers among the fish species medaka (chromosome 18) and stickleback (chromosome 7). In the pufferfish genome markers of LG3 mapped in a syntenic region, but their position were not determined (Table 3). Despite the synteny of LG3 markers identified in both species, an inversion of the markers UNH115 and UNH180 was detected in medaka in relation to *Oreochromis* (Figure 4). Using *CLCN5* gene located in the LG3 in a comparative analyzes against zebrafish genome, this gene appears duplicated and located at chromosomes 7 and 21, and in mammal species *CLCN5* has been conserved in X chromosome (Figure 4).

The majority of markers of LG5 correspond to chromosome 11 in pufferfish, chromosome 5 in medaka, and a lot of unidentified regions in stickleback genome (Table 3). The *MME* gene differs in its position appearing in chromosome 16, 13 and 1, respectively in the three species mentioned (Table 3) (Figure 5). Zebrafish was added in Ensembl datasheet and showed an extremely conserved pattern of LG5 in chromosome 11. Once more, this marker appears duplicated in zebrafish (see *c-Ski*, for instance in Figure 5). Genes observed in LG5 (17cM) seems to be much conserved in the chromosome 1 of primates, and also some conservation is related to horse chromosome 2, mouse chromosome 4, chicken chromosome 21, and turkey chromosome 23 (Figure 5). The *OPN* genes, present in the end of *O. niloticus* LG5, matched in the X chromosome for all mammals analyzed (Figure 5).

Bouillabase analyzes of LG7 markers showed a block of synteny between *Oreochromis*, pufferfish chromosome 13 and stickleback chromosome 19. In medaka genome UNH179 marker was identified in chromosome 6, but it was not possible to identify the position of the marker *IGFBP2* (Table 3, Figure 6). In Ensembl comparative study, the analyze of a small region (about 3cM) from LG7 that contain the genes *RERG*, *KCNE1L*, *DUSP6* and *IGFBP2*, reveals the conservation position at the same chromosome of *RERG* and *DUSP6*, in several vertebrates as *Homo sapiens* chromosome 12, *Macaca mulatta* chromosome 11, *Bos Taurus* chromosome 5, *Gallus gallus* and *Meleagris gallopavo* chromosome 1, and indeed in the reptile *Anolis carolinensis* chromosome 5 (Figure 6). Moreover, the gene *KCNE1L* was conserved in the X chromosome for all mammals (Figure 6).

Discussion

Cytogenetic mapping and chromosome stability in Pseudocrenilabrinae

The most important point of this work is the extreme chromosome conservation within African cichlids. It is already know that African cichlids (Pseudocrenilabrinae) contains about 1.400 species (Turner et al. 2007) and the modal chromosome number is 2n = 44 (Poletto et al. 2010). Extensive comparative mapping has demonstrated that the genetic maps for tilapia and Malawi cichlids are almost perfectly collinear (Ser et al. 2009). This conservation could be enhanced through the linkage groups used in the physical chromosomal mapping at the present work.

Within Pseudocrenilabrinae, the major feature in the karyotype of tilapiines is the presence of one large st/a chromosome pair (LG3), which is significantly larger than all others in the karyotype (Majumdar and McAndrew 1986, Poletto et al. 2010). On the

other hand, the haplochromine and hemichromine cichlids have shown two outstanding chromosome pairs (Ferreira et al. 2010; Poletto et al. 2010), the first and the second larger that correspond to LG3 and LG7, respectively. The chromosome mapping of LG3 and LG7 markers (present report) in Pseudocrenilabrinae species confirm the idea of homology between the two larger chromosomes pairs of tilapiines and haplochromines/hemichromines (Ferreira et al. 2010). However the hypothesis about independent chromosome fusion that originated the largest chromosomes in tilapiines and haplochromines suggested by Ferreira et al. (2010) is not in agreement according to the results observed for the cytogenetic mapping of LG3 and LG7 markers.

It's known that the larger chromosome pair of *O. niloticus* originated by a centric fusion event of 3 other pairs of the ancestral cichlid karyotype composed of 48 acrocentric chromosomes (Chew et al. 2002). Ferreira et al. (2010) propose that a first chromosome fusion took place before the divergence of the main East African cichlid groups. The second event of chromosome fusion occurred independently in the tilapiines and non-tilapiines. In the tilapiines, a new chromosome was fused to the largest pair, and in the non-tilapiines, the second fusion involved 2 other chromosomes giving the chromosome pair 2. However, when we compare the distribution of BAC signals (LG3 and LG7) through the long arm in tilapiines and non-tilapiines species analyzed, we show that second event of chromosome fusion should be identical in these two groups, giving chromosome with the same genomic content that these remain conserved in all species so far.

Also the differences observed in the chromosome size of the first pair among tilapiines and haplochromines should be related to the intense dynamics of repeated DNA located in the middle of the long arm instead of differences related to possible

chromosome fusion. Several studies have shown that the largest chromosome in *O. niloticus* is enriched in microsatellites, transposable elements, LINES and SINES, and non-LTR retrotransposons (Oliveira et al. 1999, Oliveira et al. 2003, Martins et al. 2004, Ferreira and Martins 2008, Ferreira et al 2010, Valente et al. 2009, 2011). The positive results using LG3-BACs (enriched of repetitive DNA, see Table 2), evidenced the accumulation of repeated DNAs in the LG3 chromosome of all Pseudocrenilabrinae, and reinforces the homology between the largest chromosome pair of all species analyzed.

Despite displaying a highly conserved karyotype structure, many events of duplication, inversion and fusion occurred during the diversification of African cichlids and resulted in diploid numbers divergent from the pattern of 44 chromosomes, as observed in *Tilapia mariae* with 2n=40 (Thompson 1981, Ferreira et al. 2010), *T. sparrmanii* with 2n=42 (Thompson 1981), *O. alcalicus* with 2n=48 (Poletto et al. 2010), *O. karongae* with 2n=38 (Mota-Velasco et al. 2010), *A. burtoni* with 2n=40 (Poletto et al. 2010), among others.

A. burtoni (2n = 40) has the presence of two metacentric chromosome pairs, which are probably the result of centric fusions of two small st/a chromosomal pairs. According to our results the ancestral st/a chromosome that contains the LG5 (observed in other haplochromines) might be involved in one of these events resulting in an m/sm chromosome. Such assumption could not be inferred for *T. mariae* (tilapiine), in which the diploid number reduction to 2n = 40 must be the result of rearrangements that did not involve the linkage groups investigated here (Figure 2).

Additionally, the reduction of chromosome number observed in *O. karongae* (2n = 38) and the presence of three pairs of medium- sized chromosomes not found in the

typical *Oreochromis* species, were originated by chromosome fusion events involving LG1 (Harvey et al. 2002, Mota-Velasco et al. 2010), and could represent recent chromosomal rearrangements that have occurred independently in tilapiine and haplochromine groups. The aditional signal observed in the 1p arm (LG3 – UNH115) of the largest chromosome of males and females of *M. lombardoi* should be a consequence of specific rearrangements or even more the presence of chimeric BAC inserts.

The absence of BAC-FISH signal in South American cichlids (Cichlinae) belonging to different tribes (Table 1), could be associated to nucleotide divergences or chromosome rearrangements that have disrupted in a small-scale level the genomic blocks carried by the BAC clones of *O. niloticus* in relation to the South American cichlins. Instead of being a monophyletic group (Sparks and Smith 2004), Neotropical cichlids harbor significantly higher levels of genetic variation compared to African Pseudocrenilabrinae group (Farias et al. 1999).

Although genomic rearrangements seem to have occurred differentiating South American and African cichlids, it is plausible that large genomic blocks are still conserved among them as well as in relation to other fish groups. Unfortunately deep analysis integrating cytogenetics and genomic data were not possible for Cichlinae because there is no large scale genomic data available for this subfamily yet.

Comparative analysis of vertebrates

There is a classical idea that fish genomes have high rates of chromosomal rearrangements compared to other vertebrates and then, probably, low rates of synteny (Robinson-Rechavi et al. 2001a, 2001b, Venkatesh 2003, Ravi and Venkatesh 2008). However, genomics studies are revealing that level of synteny among teleosts and

others vertebrates seem to be higher (Grutzner et al. 2002; Freeman et al. 2007; Putnam et al. 2007; Mazzuchelli et al. 2011, present work).

The present work reinforces that fish genomes are extremely conserved instead of highly rearranged as stated before. Using BAC-end sequence data available at Bouillabase it was possible to perform comparative analyses and detect regions of synteny between African cichlids and several model fish species and also vertebrates. Conserved (syntenic) chromosome segments have been successfully identified by means of comparative cytogenetics between cichlid markers (mainly LG5 and 7) and mammals/birds chromosomes.

Comparative cytogenetics based on BAC-FISH among different species of Salmonidae family, Atlantic salmon and rainbow trout, evidence a conservation of large syntenic blocks (Phillips et al. 2009). Study involving *SOX* genes region in cichlids reveal a large genomic block that was conserved through vertebrates (Mazzuchelli et al. 2011). The conservation of gene order among fishes at scales of several Mb allows the use of relatively complete sequences of model fish species to accelerate gene discovery and positional cloning of non-model species (Lee and Kocher 2007a; Soler et al. 2010).

Comparative analysis of LG1 markers that covers about 11 Mb of cichlid genomes (markers GM201/UNH995) with pufferfish genome evidences a conservation of an extensive genome block between the two species (2,7~4.8 Mb) (present work, Lee and Kocher 2007a). Comparative BAC mapping has already shown that some stickleback chromosomes have a nearly complete synteny with those of cichlids. Stickleback does not belong to the order Perciformes, but is the most closely related genome to cichlids among the currently available sequenced teleost genomes (Soler et al. 2010) and is currently the best reference sequence for building comparative maps of tilapia (Sarrapoulou et al. 2008).

LG3 have a lot of markers; however it contains few known genes, which could be used for comparative analyzes in other vertebrates. *CLCN5*, a gene associated with Dent's disease disorder, a renal tubular dysfunction, was the only gene identified in other vertebrates and in mammals *CLCN5* is mapped in chromosome X (Levtchenko et al. 2007).

Several genes from different LGs of cichlids are located in human and other mammals X chromosomes (*CLCN5, OPSIN, KCNE1L*) highlighting the conservation of X chromosome through mammals (Murphy et al. 1999). However in bird clade the genes observed in mammals X chromosome does not correspond to Z or W chromosomes, instead they are located in autosomes. This is explained by the fact that sexual chromosomes ZW and XY have no homology and both sexual chromosome pair systems were derived from different autosomes from their common ancestor (Kawai et al. 2007).

In contrast to mammals and birds, the pufferfish, like most fish, does not possess heteromorphic sex chromosomes (Grützner et al. 1999), and the genetic mechanism(s) of sex determination is still unclear. At the same time, pufferfish genome revealed that the human X is an orthologous mosaic of three chromosomes: most human Xp and Xq genes are syntenic on pufferfish chromosome 1, 2 and 7 (Grützner et al. 2002). In addition, zebrafish LGs 9 and 23 are also related with both human Xp and Xq orthologs (Barbazuk et al. 2000, Woods et al. 2000).

The largest chromosome of Nile tilapia that contains the LG3 was previously thought to be the sex chromosome for Nile tilapia (Foresti et al. 1993, Carrasco et al. 1999, Harvey et al. 2003, Ocalewicz et al. 2009), however the major sex-determining region in the Nile tilapia was mapped in LG1 (Lee et al. 2003) located in a small st/a chromosome (present work, Cnaani et al. 2008). LG1 also contain *WT1* and *CYP19a*

genes, involved in mammalian sex differentiation and instead of being considered candidate genes for sex determination in vertebrates in which major sex determining genes have not been identified (D'Cotta et al. 2001; Hossain and Saunders 2001; Kwon et al. 2001; Chang et al. 2005, Lee and Kocher 2007b), they were not detected in any sexual chromosome of vertebrates analyzed here.

Fish genome duplication is a problem for comparative mapping and it has been argued that the abundance of duplicate genes in teleost might be due to independent gene duplications in different lineages rather than to whole genome duplication (Robinson-Rechavi et al. 2001a, 2001b). *WT1* is duplicated in *O. niloticus* genome (*WT1a* and *WT1b* genes) and belongs to two different tilapia BAC contigs located at LG7 and LG1, respectively. However *WT1b* shows higher coding-sequence similarity to the human *WT1* than does *WT1a* (Lee and Kocher 2007b). In *D. rerio WT1b* is located in chromosome 18, the same that contains other markers from *O. niloticus* LG1. Additionally the *WT1a* appears in chromosome 25, which also contains a marker from *O. niloticus* LG7 (*DUSP6*). However, some markers appear duplicated in *D. rerio* genome: *CLCN5* – LG3 in chromosomes 21 and 7, and *c-Ski* – LG5 in chromosomes 8 and 11, but are not duplicated in *O. niloticus* genome. Also several markers used in this work match to a single region in zebrafish genome (*CYP19a*, *AKR*, *MME*, *OPSIN*, *ATP*, *RERG*, *DUSP6*) concluding that duplicated genes are evolving, changing their nucleotide sequences, or even more the second copy may have been lost during teleost evolution.

A range of studies using comparative cytogenetics has demonstrated that karyotypes are conserved trough vertebrates. For example, despite the extensive variation in chromosome number and morphology of Canidae, the majority of conserved chromosome segments appear to have remained largely intact in the

karyotypes of extant canid species, although the relative orientation is not always conserved (Becker et al. 2011). Conserved karyotypes are observed either in birds, which show a slow rate of interchromosomal rearrangements, with most species showing diploid numbers between 76 and 80 (reviewed in Ellegren 2010). Giovannotti et al. (2009) recently showed highly conserved karyotypes within a reptile family (Scincidae) by cross-species chromosome painting. At different taxonomic level, comparative gene mapping has revealed highly conserved linkage homology between an agamid lizard (*Leiolepis reevesii*) and a snake (*Elaphe quadrivirgata*) (Srikulnath et al. 2009). This conservation of synteny spread in different vertebrates could be due to intrinsic chromosomal properties that confer selective pressure on large genomic blocks to conserve its synteny (Grutzner et al. 2002). This situation confirms the general rule of high evolutionary genome conservation as first proposed on the basis of chromosome banding in mammals up to current comparisons of genome assemblies (Graphodasky 1989; O'Brien and Menninger 2006; Graphodasky et al. 2011).

Conclusions

African cichlids have a common ancestor; share genes related to sex determination and have a conserved karyotype structure with the presence of a large chromosome pair that contains the LG3. This chromosome becoming an excellent marker for this group, and the use of BACs containing genes/markers represents a promising alternative for a better physical mapping for cichlids. Progress in methodologies that integrate genomic data and cytological analysis appears to be very significant for understanding various issues such as chromosome and genome evolution. These results evidence that large syntenic chromosome segments are evolving together in all groups of vertebrates including cichlid fish, demystifying that fishes have high

rates of chromosomal rearrangements and probably low rates of synteny. Sequence comparisons show that these regions should be part of important linkage groups because are strongly conserved between species of different groups. Comparative analysis of fish genomes in relation to other vertebrates will allow inferring on the evolution of genes in a higher scale order, contributing in the establishment of a framework for additional genome-wide studies.

List of Abbreviations

AKR: aldo-keto reductase

ATP: ATPase, Na+/K+ transporting

BAC: bacterial artificial chromosomes

CLCN5: chloride channel 5

c-Ski: v-ski sarcoma viral oncogene homolog

CYP19A1: cytochrome P450, family 19, subfamily A, polypeptide 1

DUSP6: dual specificity phosphatase 6

ESTs: expressed sequence tags

FISH: fluorescence *in situ* hybridization

IGFBP2: insulin-like growth factor binding protein 2

KCNE1L: potassium voltage-gated channel, Isk-related family, member 1-like

LG: linkage groups

m/sm: meta/submetacentric

MME: membrane metallo-endopeptidase

OPN: opsin – cone pigments

RERG: RAS-like, estrogen-regulated, growth inhibitor

SOX: sex determining region Y

st/a: subtelo/acrocentric

WT1b: Wilms tumor 1

Author's contributions

JM carried out chromosome preparation, molecular cytogenetics experiment with BAC probes, *in silico* analyses and interpretation of data, and drafted the manuscript. FY conceived the initial chromosome hybridization experiments with BAC probes, analysis and interpretation of data and revised the manuscript. TDK helps in the analysis and interpretation of data and revised the manuscript. CM conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This study was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) from Brazil. The authors are grateful to Beyone Fu – Sanger Institute for initial help on BAC-FISH.

Figure captions

Figure 1: Cytogenetic mapping based on FISH of *O. niloticus* BAC clones in *O. niloticus* metaphasic chromosomes. (A) Co-hybridization of three differentially labeled LG1 BAC-clones. (B) Co-hybridization of four differentially labeled LG3 clones to the largest chromosome. Double-fish of markers of (C) LG5 and (D) LG7 BACs. The chromosomes are counterstained with DAPI and the markers hybridized are indicated in different colors. Scale bar 10μm.

Figure 2: Chromosomal homologies revealed by BAC-FISH analyze. This figure summarizes the results of the comparative FISH mapping of markers of LG1, 3, 5 and 7 in nine Pseudocrenilabrinae species. For more details see Additional file 2. The different colors indicate the markers mapped and the asterisks (*) indicate markers that carry repeated DNA sequences.

Figure 3: Comparative analyzes of LG1 showing the conservation of genomic blocks between Nile tilapia, medaka, stickleback and pufferfish, and other vertebrates. BAC-labeled metaphasic chromosomes are showed on the left, followed by ideograms and LG1 of *O. niloticus*. The three markers (*CYP19a*, *WT1* and UNH995) are identified by different colors. On the right, a summary of the results of comparative analyzes with other vertebrates. The numbers represent the chromosome identified and the conserved genomic regions are highlighted in different colors.

Figure 4: Comparative analyzes of LG3 showing the conservation of genomic blocks between Nile tilapia, medaka, stickleback and pufferfish and other vertebrates. BAC-labeled metaphasic chromosomes are showed on the left, followed by ideograms and LG3 of *O. niloticus*. The ideogram was inverted according to LG3 orientation. The markers (TRP1, CLCN5, UNH115, GM180, GM526, UNH106, GM354, GM204) are identified by different colors. On the right, a summary of the results of comparative analyzes with other vertebrates. The numbers represent the chromosome identified and the conserved genomic regions are highlighted in different colors.

Figure 5: Comparative analyzes of LG5 showing the conservation of genomic blocks between Nile tilapia, medaka, stickleback and pufferfish and vertebrates. BAC-labeled metaphasic chromosomes are showed on the left, followed by ideograms and LG5 of *O. niloticus*. The ideogram was inverted according to LG5 orientation. The four markers (*Blue/red OPSIN, Green OPSIN, c-SKI, ATP*) are identified by different colors. Below, a summary of the results of comparative analyzes with other vertebrates. The numbers represent the chromosome identified and the conserved genomic regions are highlighted in different colors.

Figure 6: Comparative analyzes of LG7 showing the conservation of genomic blocks between Nile tilapia, medaka, stickleback and pufferfish and other vertebrates. BAC-labeled metaphasic chromosomes are showed on the left, followed by ideograms and LG7 of *O. niloticus*. The ideogram was inverted according to LG7 orientation. The three markers (UNH179, *DUSP6*, UNH896) are identified by different colors. Below, a summary of the results of comparative analyzes with other vertebrates. The numbers represent the chromosome identified and the conserved genomic regions are highlighted in different colors.

Tables

Table 1: Cichlidae species submitted to cytogenetic mapping using BAC clones of *O. niloticus* as probes in FISH experiments. na, non available data.

Subfamily	**Groups or Tribes**	**Species**	**2n**	**Sex**	**Karyotype**	**Origin**
Etroplinae		*Etroplus maculatus*	46	1M/1F/1na	18m/sm+18st/a+10 micro	Petshop
Pseudocreilabinae	Tilapiinae	*Oreochromis niloticus*	44	2M/3F/1na	2m/sm+42st/a	TAF-UMD
		Oreochromis mossambicus	44	1M/1F	4m/sm+40st/a	TAF-UMD
		Oreochromis aureus	44	1F	2m/sm+42st/a	TAF-UMD
		Tilapia mariae	40	1M/1F	8m/sm+32st/a	TAF-UMD
	Haplochrominae	*Astatotilapia latifasciata*	44	2M/3F/1na	12m/sm+32st/a	Petshop
		Metriaclima lombardoi	44	1M/3F/1na	14m/sm+30st/a	TAF-UMD
		Astatotilapia burtoni	40	1M/2F	14m/sm+26st/a	TAF-UMD
		Labeotropheus trewavase	44	2M/3F	14m/sm+30st/a	TAF-UMD
	Hemichrominae	*Hemichromis bimaculatus*	44	?	4m/sm+40st/a	Petshop
Cichlinae	Cichlini	*Cichla kelberi*	48	2M	48 st/a	Araguaia river
	Astronotini	*Astronotus ocellatus*	48	2M	12m/sm+36st/a	Tietê river
	Heroini	*Symphysodon aequifasciatus*	60	3na	46m/sm+4st/a+10micro	Petshop
	Geophagini	*Geophagus brasiliensis*	48	1M+1F	2m/sm+46st/a	SP State rivers

2n= diploid number; m/sm= metacentric/submetacentric; st/a= subtelocentric/acrocentric; M/F= males or/and females

Table 2: Genetic markers and their BAC identification (ID)

Markers	**LG**	**BAC 384 Well ID**	**Gene**	**Observations**
UNH995	LG1	b04TI071H11		Cnaani et al 2008[+]
	LG1	b04TI008J05	*CYP19A*	
	LG1	b04TI002B08	*CYP19A*	
	LG1	b03TI091I08	*WT1*	
GM354	LG3	b03TI066P02		Cnaani et al 2008[+]
GM204	LG3	b04TI071O04		Cnaani et al 2008[+]
UNH180	LG3	b04TI056G07		Cnaani et al 2008[+]
UNH115	LG3	b03TI086K09		Cnaani et al 2008[+]
	LG3	b04TI076F11	*CLCN5*	Cnaani et al 2008[+]
	LG3	b03TI073M01	*TRP1*	
UNH106	LG3	b03TI088C14		
GM526	LG3	b03TI067N14		
	LG5	b04TI053F24	*ATP*	3.65 Mb Tetr 11
	LG5	b04TI010O22	*Green OPSIN*[+]	
	LG5	b04TI075I09	*Blue/red OPSIN*	
	LG5	b04TI006L21	*c-SKI*	
	LG7	b03TI050E01	*RERG*	
	LG7	b03TI079D23	*KCNE1L*[+]	
	LG7	b03TI080A15	*DUSP6*	
	LG7	b03TI081O07	*IGFBP2*	
UNH179	LG7	b04TI036P14		
UNH896	LG7	b04TI035B08		

[+]: Presence of repetitive DNAs

Table 3: Bouillabase comparative analyze using BAC data from *Oreochromis niloticus* against three fish models. NID, not identified scaffold in *O. niloticus*; ns, no similarity; chrUn, unidentified chromosome.

LG	Marker/gene	Scaffold	medaka	stickleback	pufferfish
1	*CYP19A1*	287	chr3:12108168..12108464	ns	chr5:9794344..9794562
1	ACG/CTT382	17	chr3:20997422..20997535	chr2:12085900..12086028	chr5:4855030..4855202
1	BJ690985	NID	chr19:4811402..4811495	ns	ns
1	*WT1*	NID	chr3:22126505..22127094	chr2:12884672..12885323	chr5:4189140..4189790
3	UNH180	143	chr18:20335074..20335221	ns	chrUn:92504595..92504738
3	UNH115	NID	chr18:12554314..12554397	chrUn:7877924..7878007	chrUn:98704287..98704370
3	*TRP1*	88	chrUn:35101810..35101948	chr7:5614225..5614569	chrUn:56261613..56261704
3	GM526	89	chr18:27239660..27426308	chr7:10064534..10064979	ns
5	*AKR*	155	chr5:7748231..7748360	chrUn:4859687..4859812	chr11:3868317..3868442
5	*AKR*	155	chr5:7708974..7709058	chrUn:4889317..4982133	chr11:3884393..3884499
5	*MME*	340	chr13:11994389..11995120	chr1:10620278..10620987	chr16:2528394..2528542
5	*MME*	162	chr13:12106428..12106606	chr1:10575763..10575935	ns
5	*Green OPSIN*	NID	chrUn:142344721..142344800	chrUn:4181024..4181104	chr11:5142370..5142445
5	*Blue/red OPSIN*	19	chr5:26959747..26959838	chr17:10579335..10579447	ns
5	*c-SKI*	NID	chr5:6600759..6600879	chrUn:3593246..3593357	chr11:4411020..4411130
5	*c-SKI*	64	chr5:6741965..6742076	chrUn:3505497..3505978	chr11:4336173..4336280
7	*DUSP6*	52	ns	chr19:13,105,072..13,105,575	chr13:6,253,123..6,253,497
7	*IGFBP2*	NID	chrUn:75,506,034..75,506,204	chrUn:62,550,041..62,550,211	chr13:6,057,596..6,145,055
7	UNH179	NID	chr6:14,100,033..14,100,152	chr19:16,793,012..16,793,135	chr13:3,161,846..3,161,931
7	UNH896	NID	ns	chr19:2,835,715..2,992,804	ns

Figure 1

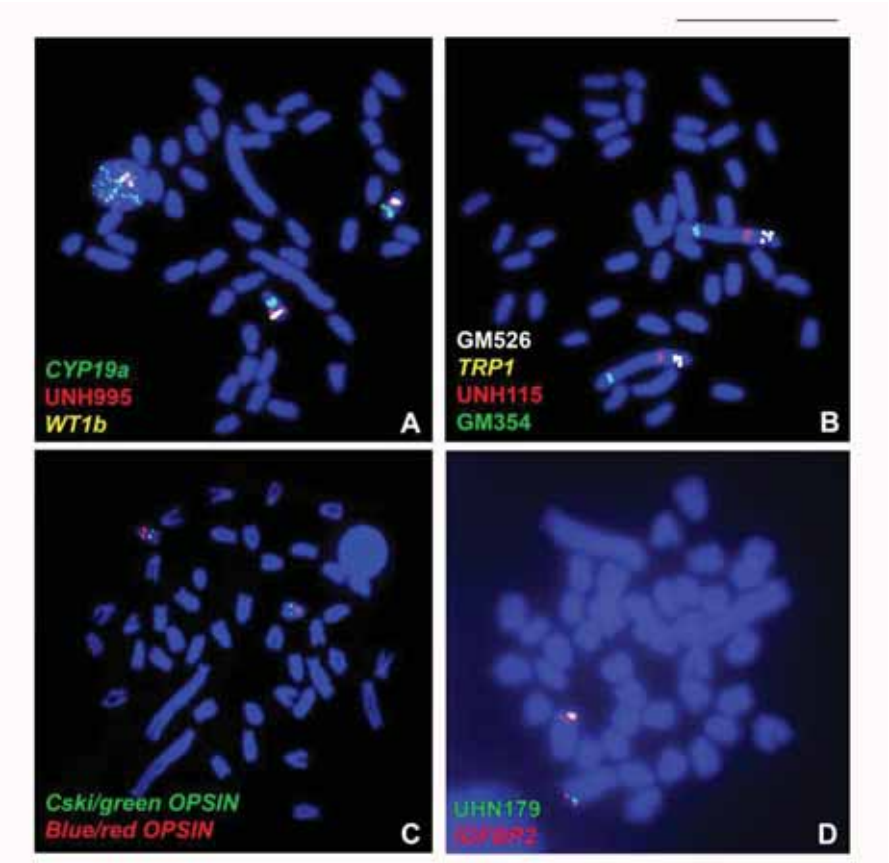

CYP19a
UNH995
WT1b
A
GM526
TRP1
UNH115
GM354
B
Cski/green OPSIN
Blue/red OPSIN
C
UHN179
D

Figure 2

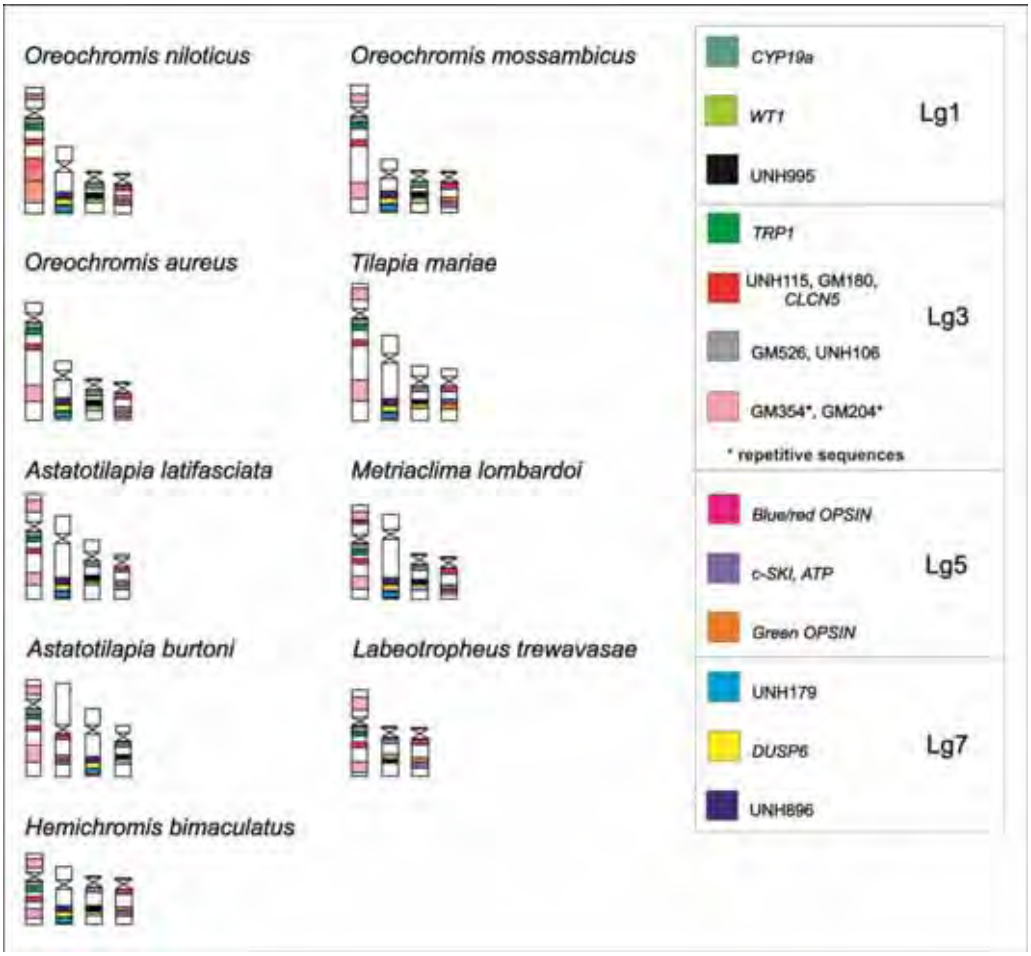

Oreochromis niloticus
Oreochromis mossambicus
Oreochromis aureus
Tilapia mariae
Astatotilapia latifasciata
Metriaclima lombardoi
Astatotilapia burtoni
Labeotropheus trewavasae
Hemichromis bimaculatus
CYP19a
WT1
UNH995
Lg1
TRP1
UNH115, GM180, CLCN5
GM526, UNH106
GM354*, GM204*
Lg3
* repetitive sequences
Blue/red OPSIN
c-SKI, ATP
Green OPSIN
Lg5
UNH179
DUSP6
UNH896
Lg7

Figure 3

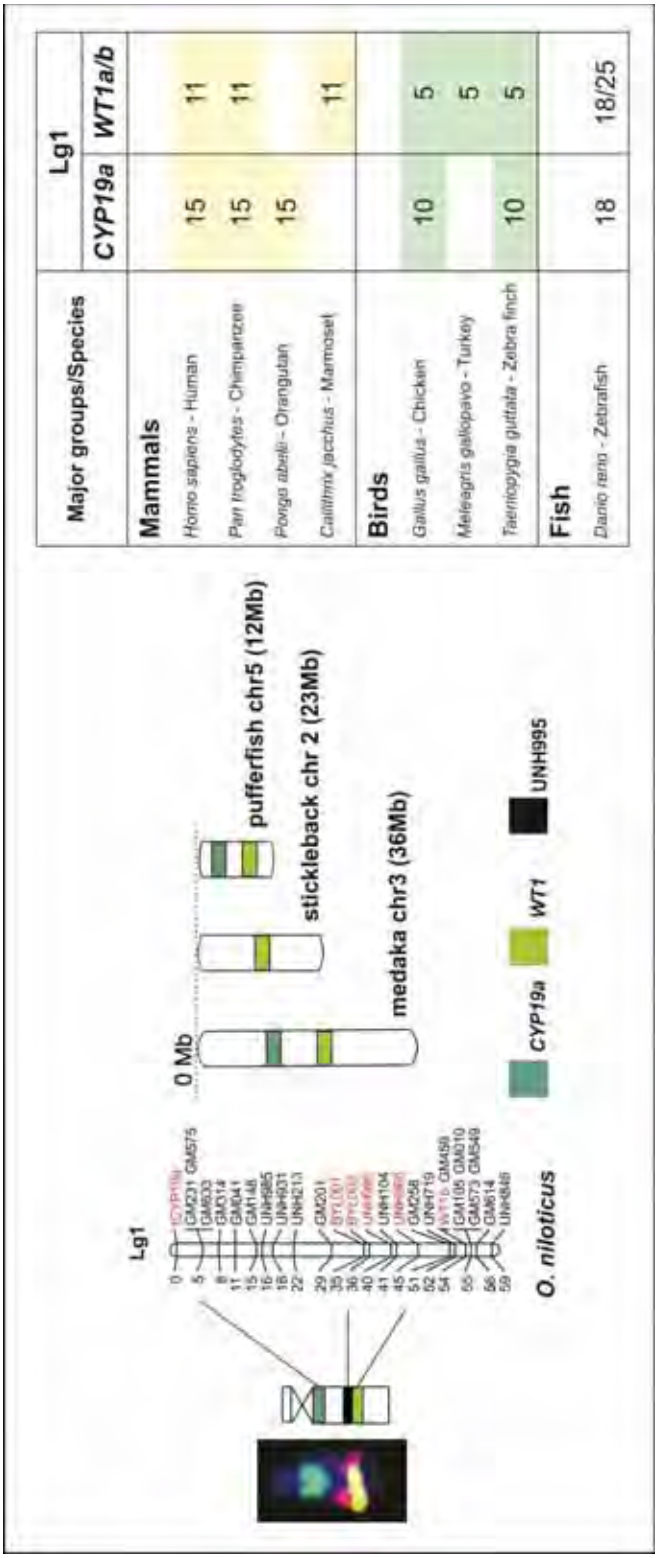

Major groups/Species
Lg1
CYP19a
WT1a/b
Mammals
Homo sapiens - Human
15
11
Pan troglodytes - Chimpanzee
15
11
Pongo abelii - Orangutan
15
Callithrix jacchus - Marmoset
11
Birds
Gallus gallus - Chicken
10
5
Meleagris gallopavo - Turkey
5
Taeniopygia guttata - Zebra finch
10
5
Fish
Danio rerio - Zebrafish
18
18/25
Lg1
0 Mb
pufferfish chr5 (12Mb)
stickleback chr 2 (23Mb)
medaka chr3 (36Mb)
CYP19a
WT1
UNH995
O. niloticus

Figure 4

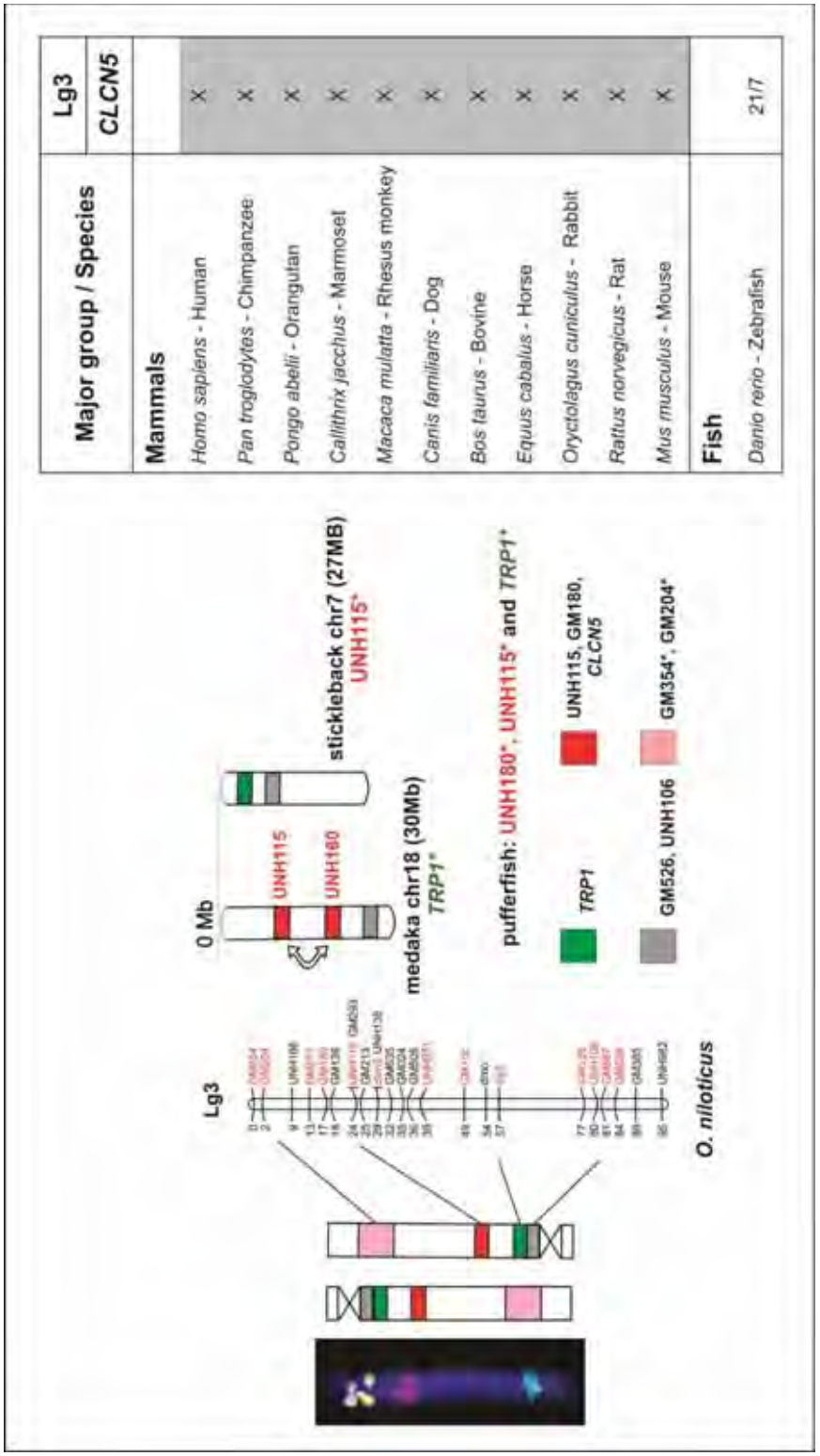
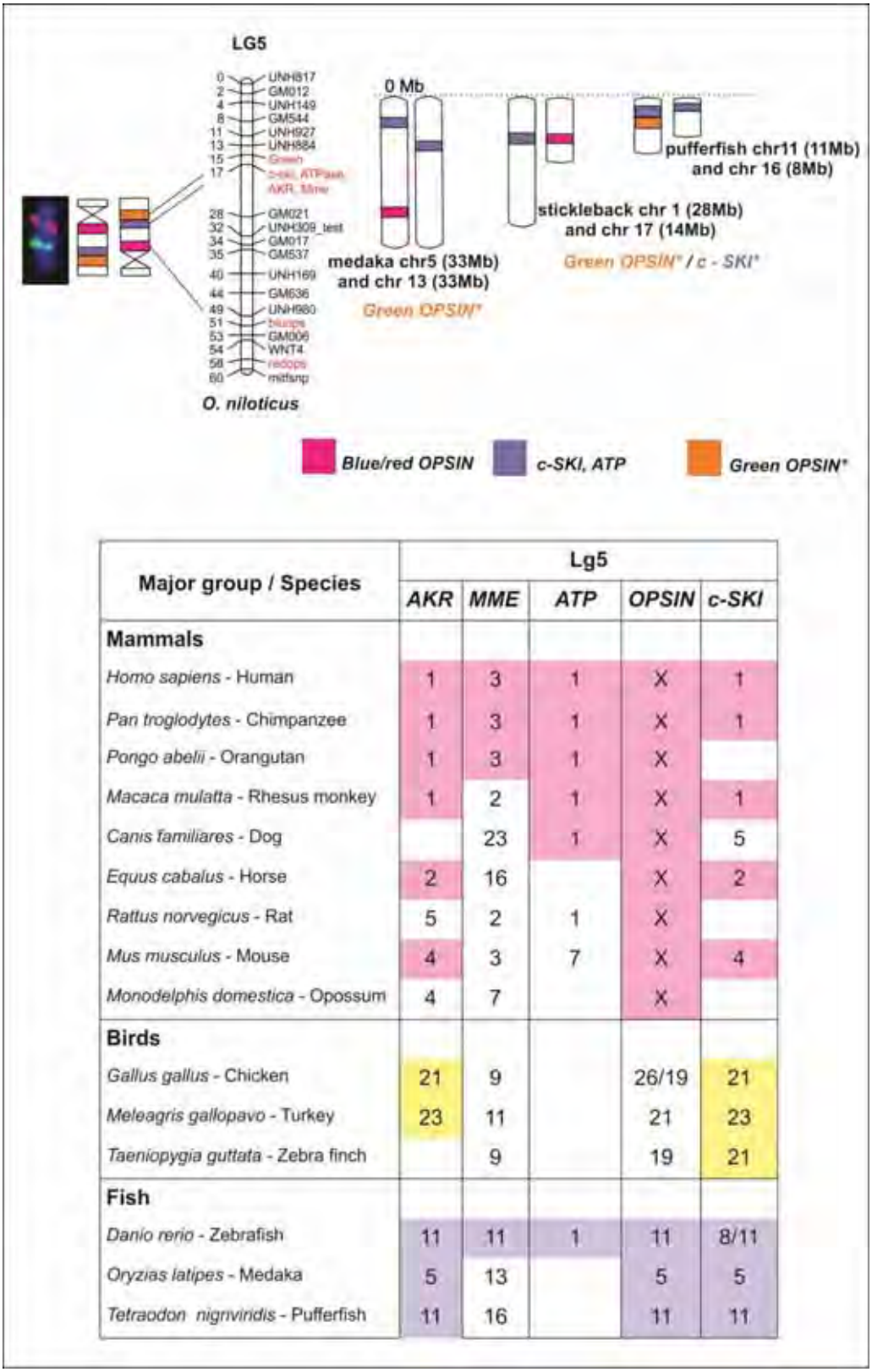
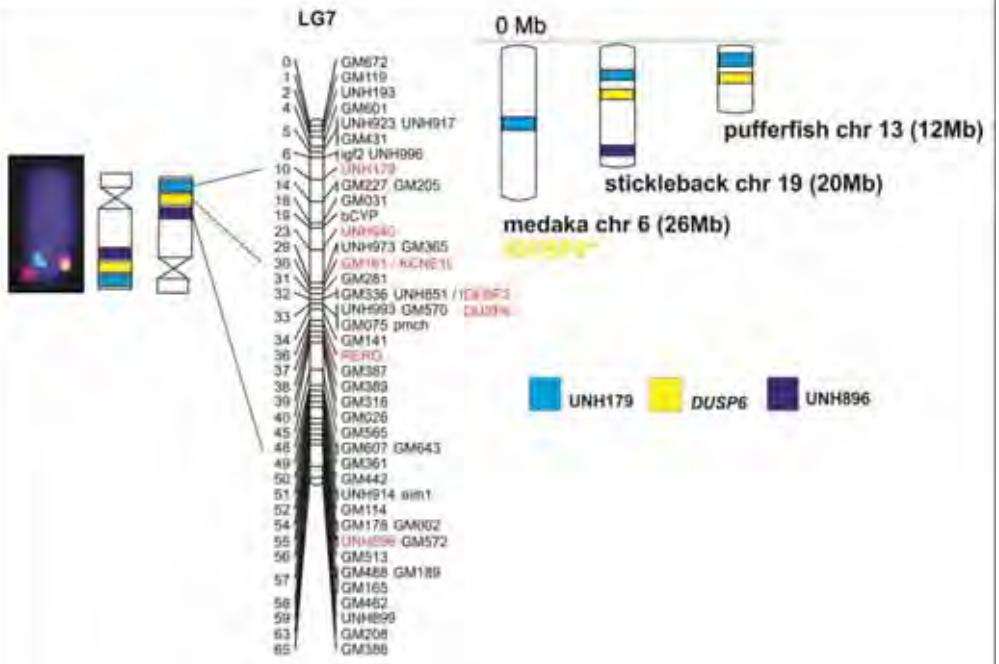

Major group / Species
Lg3
CLCN5
Mammals
Homo sapiens - Human
Pan troglodytes - Chimpanzee
Pongo abelii - Orangutan
Callithrix jacchus - Marmoset
Macaca mulatta - Rhesus monkey
Canis familiaris - Dog
Bos taurus - Bovine
Equus cabalus - Horse
Oryctolagus cuniculus - Rabbit
Rattus norvegicus - Rat
Mus musculus - Mouse
Fish
Danio rerio - Zebrafish
21/7
0 Mb
UNH115
UNH180
stickleback chr7 (27MB)
UNH115*
medaka chr18 (30Mb)
TRP1*
pufferfish: UNH180*, UNH115* and TRP1*
UNH115, GM180, CLCN5
TRP1
GM354*, GM204*
GM526, UNH106
Lg3
O. niloticus

Figure 5



Major group / Species	Lg5				
	AKR	*MME*	*ATP*	*OPSIN*	*c-SKI*
Mammals					
Homo sapiens - Human	1	3	1	X	1
Pan troglodytes - Chimpanzee	1	3	1	X	1
Pongo abelii - Orangutan	1	3	1	X	
Macaca mulatta - Rhesus monkey	1	2	1	X	1
Canis familiares - Dog		23	1	X	5
Equus cabalus - Horse	2	16		X	2
Rattus norvegicus - Rat	5	2	1	X	
Mus musculus - Mouse	4	3	7	X	4
Monodelphis domestica - Opossum	4	7		X	
Birds					
Gallus gallus - Chicken	21	9		26/19	21
Meleagris gallopavo - Turkey	23	11		21	23
Taeniopygia guttata - Zebra finch		9		19	21
Fish					
Danio rerio - Zebrafish	11	11	1	11	8/11
Oryzias latipes - Medaka	5	13		5	5
Tetraodon nigriviridis - Pufferfish	11	16		11	11

Figure 6



Major groups / Species	Lg7			
	RERG	*KCNE1L*	*DUSP6*	*IGFBP2*
Mammals				
Homo sapiens - Human	12	X	12	2
Pan troglodytes - Chimpanzee	12		12	
Pongo abelii - Orangutan	12	X	12	
Calithrix jacchus - Marmoset	9	X	9	
Macaca mulatta - Rhesus monkey	11	X	11	12
Canis familiaris - Dog	27	X	15	37
Bos taurus - Bovine	5	X	5	2
Sus scrofa - Pig	5	X		
Oryctolagus cuniculus - Rabbit	8	X	4	7
Mus musculus - Mouse	6	X	10	1
Monodelphis domestica - Opossum	8		8	7
Birds				
Gallus gallus - Chicken	1	4	1	7
Meleagris gallopavo - Turkey	1		1	7
Taeniopygia guttata - Zebra finch				7
Reptiles				
Anolis carolinensis - Lizard	5		5	

Additional files

File name: Additional File 1
File format: pdf Adobe Acrobater
Title of data: Genomicus/Ensembl accession number
Description of data: Datasheet with accession numbers and markers of all sequences analyzed in Genomicus/Ensembl database for the LG1, 3, 5 and 7. Analyzes were conducted in July 2011.

File name: Additional File 2
File format: pdf Adobe Acrobater
Title of data: Metaphase spreads of cichlid species probed with BAC clones from *Oreochromis niloticus*.
Description of data: BACs containing markers of LG 1, 3, 5 and 7 were hybridized under FISH procedure and are indicated in different colors in each metaphase. The arrows indicate the chromosome position of probes.

Additional file 1:

Datasheet with accession numbers and markers of all sequences analyzed in Genomicus/Ensembl database for the LG1, 3, 5 and 7.

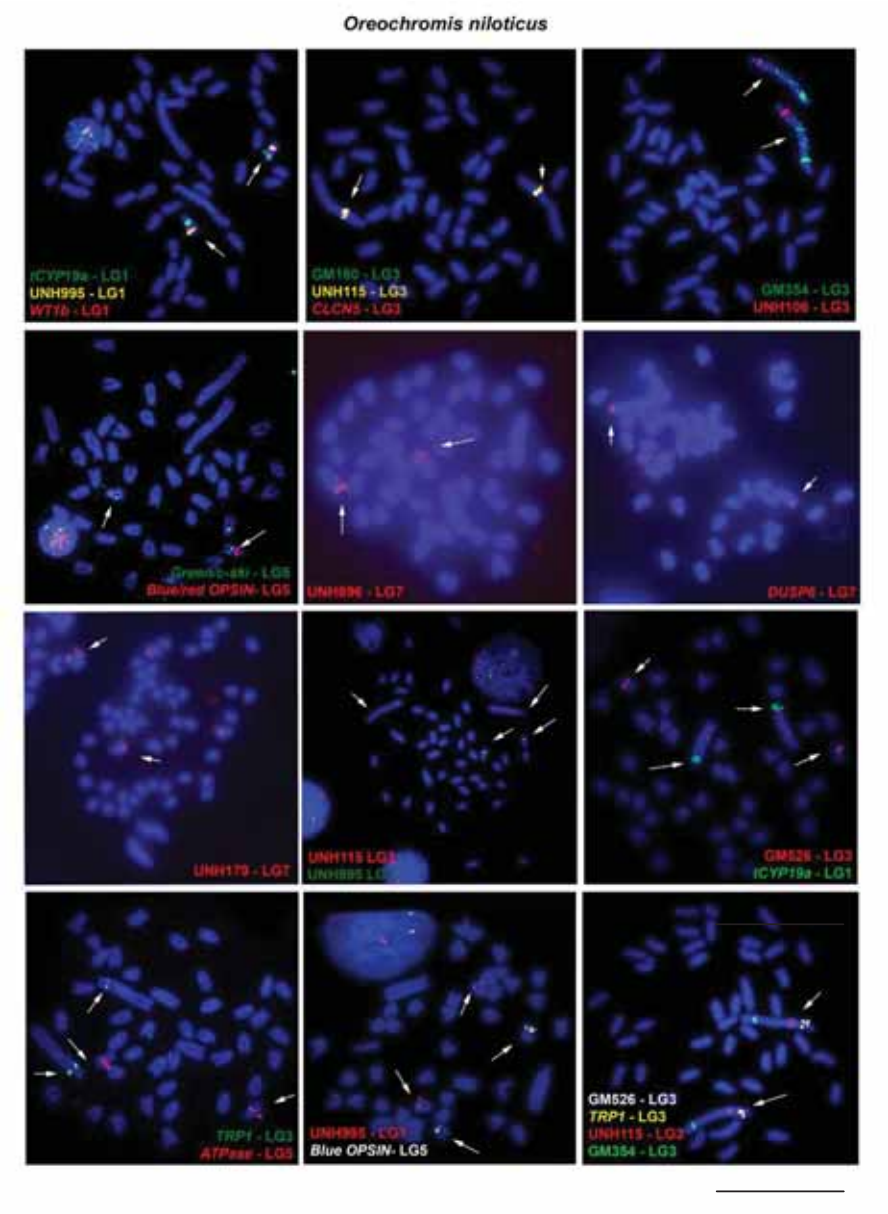
Analyzes were conducted in July 2011.

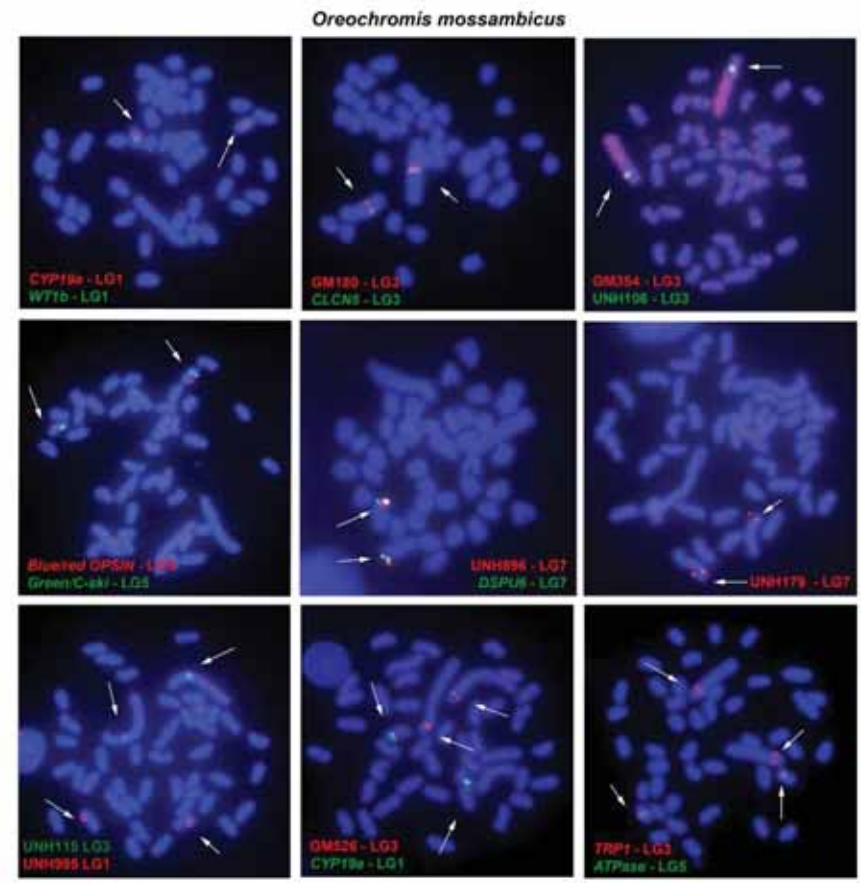
Major group/Species	Markers of LG1		Markers of LG3
	CYP19a	Wt1a/b	CLCN5
Mammals			
Homo sapiens - Human	ENSG00000137869	ENSG00000184937	ENSG00000171365
Pan troglodytes - Chimpanzee	ENSPTRG00000007074	ENSPTRG00000003476	ENSPTRG00000021899
Pongo abelii - Orangutan	ENSPPYG00000006475		ENSPPYG00000020349
Callithrix jacchus - Marmoset		ENSCJAG00000005900	ENSCJAG00000002945
Macaca mulatta – Rhesus monkey	ENSMMUG00000002553	ENSMMUG00000009881	ENSMMUG00000023056
Canis familiaris - Dog	ENSCAFG00000015338	ENSCAFG0000000742	ENSCAFG00000015971
Bos taurus - Bovine	ENSBTAG00000014890		ENSBTAG00000001507
Equus cabalus - Horse	ENSECAG00000020474	ENSECAG00000022308	ENSECAG00000010745
Sus scrofa - Pig		ENSSSCG00000013316	
Oryctolagus cuniculus - Rabbit	ENSOCUG00000012303	ENSOCUG00000005882	ENSOCUG00000017336
Rattus norvegicus – Rat	ENSRNOG00000000196	ENSRNOG00000013074	ENSRNOG00000002862
Mus musculus – Mouse		ENSMUSG00000016458	ENSMUSG00000004317
Monodelphis domestica - Opossum		ENSMODG00000009646	ENSMODG00000011303
Birds			
Gallus gallus – Chicken	ENSGALG00000013294	ENSGALG00000012115	ENSGALG00000007234
Meleagris gallopavo - Turkey		ENSMGAG00000001489	ENSMGAG00000004934
Taeniopygia guttata – Zebra finch	ENSTGUG00000006993	ENSTGUG00000004906	ENSTGUG00000005324
Reptiles			
Anolis carolinensis - Lizard		ENSACAG00000002480	
Fish			
Danio rerio – Zebrafish	ENSDARG00000041348	ENSDARG00000007990 / ENSDARG00000031420	ENSDARG00000019693 / ENSDARG00000022466
Oryzias latipes - Medaka	ENSORLG00000002949		ENSORLG00000008602
Tetraodon nigroviridis - Pufferfish	ENSTNIG00000009520		
Gasterosteus aculeatus – Stickleback			

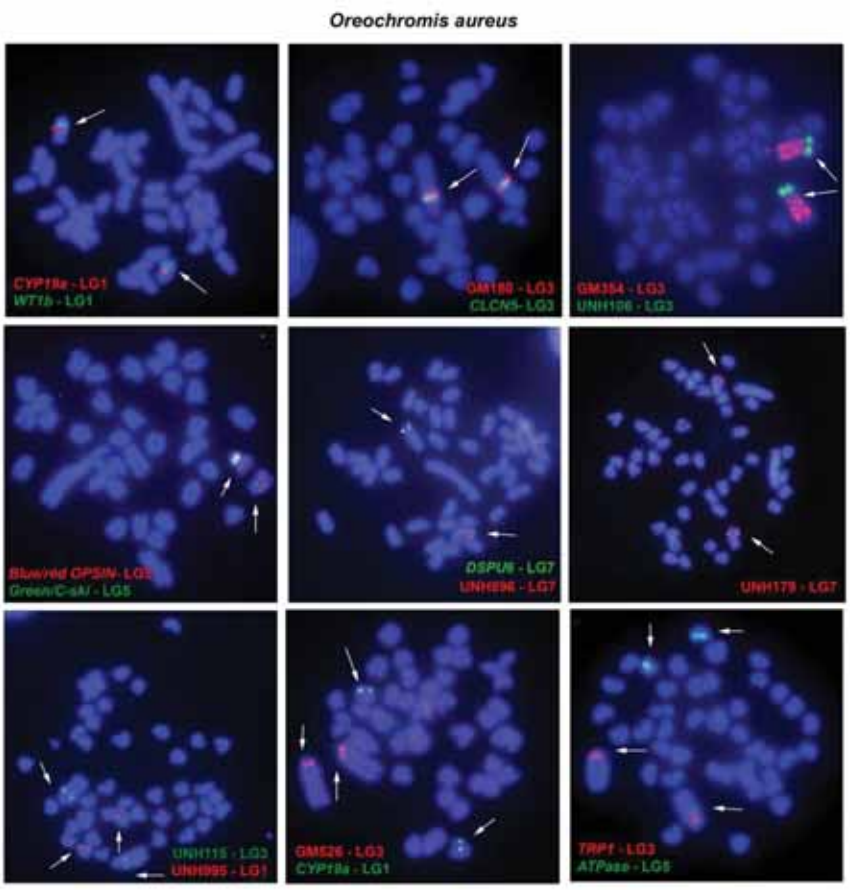
Major group/Species	Markers of LG5				
	AKR	Mme	ATPase	Opsin	C-ski
Mammals					
Homo sapiens - Human	ENSG00000117448	ENSG00000196549	NM_001160234.1	ENSG00000102076	ENSG00000157933
Pan troglodytes - Chimpanzee	ENSPTRG00000000676	ENSPTRG00000015550	XM_513679.3	ENSPTRG00000022427	ENSPTRG00000000047
Pongo abelii - Orangutan	ENSPPYG00000001410	ENSPPYG00000014224	NM_001133855.1	ENSPPYG00000020867	
Callithrix jacchus - Marmoset		ENSCJAG00000000959	XM_002751283.1		
Macaca mulatta – Rhesus monkey	ENSMMUG00000020085	ENSMMUG00000006662	XM_001112641.2	ENSMMUG00000006166	ENSMMUG00000010280
Canis familiaris - Dog		ENSCAFG00000008711	XM_850193.1	ENSCAFG00000019441	ENSCAFG00000019387
Bos taurus - Bovine	ENSECAG00000000223	ENSBTAG00000002075		ENSBTAG00000020299	ENSBTAG00000038716
Equus cabalus - Horse	ENSECAG00000006082	ENSECAG00000009138		ENSECAG00000018684	ENSECAG00000012543
Sus scrofa - Pig	ENSSSCG00000003491	ENSSSCG00000011723			
Oryctolagus cuniculus - Rabbit	ENSOCUG00000014622	ENSOCUG00000017924			
Rattus norvegicus – Rat	ENSRNOG00000017780	ENSRNOG00000009514	NM_012506.1	ENSRNOG00000037261	
Mus musculus – Mouse	ENSMUSG00000028743	ENSMUSG00000027820	BC037206.1	ENSMUSG00000031394	ENSMUSG00000029050
Monodelphis domestica - Opossum	ENSMODG00000015881	ENSMODG00000015805		ENSMODG00000025355	
Birds					
Gallus gallus – Chicken	ENSGALG00000004005	ENSGALG00000010331		ENSGALG00000002848/19	ENSGALG00000001229
Meleagris gallopavo - Turkey	ENSMGAG00000005392	ENSMGAG00000010779		ENSMGAG00000007057	ENSMGAG00000002204
Taeniopygia guttata – Zebra finch		ENSTGUG00000011304		ENSTGUG00000007365	ENSTGUG00000002944
Reptiles					
Anolis carolinensis - Lizard		ENSACAG00000003903		ENSACAG00000012605	
Fish					
Danio rerio - Zebrafish	ENSDARG00000016649	ENSDARG00000026398	NM_131687.1	ENSDARG00000044862	ENSDARG00000042151/ ENSDARG00000008034
Oryzias latipes - Medaka	ENSORLG00000004388	ENSORLG00000005790		ENSORLG00000014058	ENSORLG00000003949
Tetraodon nigroviridis - Pufferfish	ENSTNIG00000014801	ENSTNIG00000007690		ENSTNIG00000007322	ENSTNIG00000014776
Gasterosteus aculeatus – Stickleback		ENSGACG00000010305			

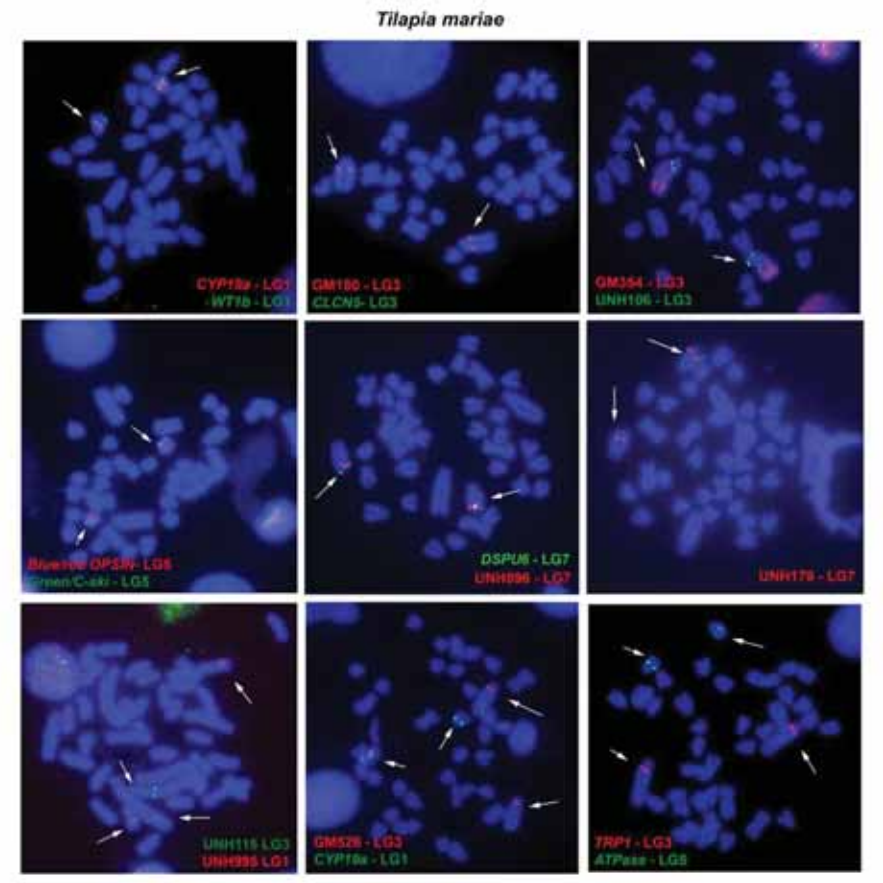
Major group/Species	Markers of LG7			
	RERG	**KCNE1L**	**DUSP6**	**IGFBP2**
Mammals				
Homo sapiens – Human	ENSG00000134533	ENSG00000176076	ENSG00000139318	ENSG00000115457
Pan troglodytes – Chimpanzee	ENSPTRG00000004731		ENSPTRG00000005282	
Pongo abelii – Orangutan	ENSPPYG00000004331	ENSPPYG00000020626	ENSPPYG00000004817	
Callithrix jacchus – Marmoset	ENSCJAG00000021602	ENSCJAG00000023434	ENSCJAG00000019653	
Macaca mulatta – Rhesus monkey	ENSMMUG00000010460	ENSMMUG00000004187	ENSMMUG00000006227	ENSMMUG00000018940
Canis familiaris – Dog	ENSCAFG00000012854	ENSCAFG00000018048	ENSCAFG00000006100	ENSCAFG00000014490
Bos taurus – Bovine	ENSBTAG00000002174	ENSBTAG00000012319	ENSBTAG00000004587	ENSBTAG00000005596
Equus cabalus – Horse	ENSECAG00000014358		ENSECAG00000010089	ENSECAG00000012058
Sus scrofa – Pig	ENSSSCG00000000602	ENSSSCG00000012582		
Oryctolagus cuniculus – Rabbit	ENSOCUG00000000650	ENSOCUG00000005007	ENSOCUG00000002902	ENSOCUG00000008561
Rattus norvegicus – Rat	ENSRNOG00000027592		ENSRNOG00000023896	ENSRNOG00000016957
Mus musculus – Mouse	ENSMUSG00000030222	ENSMUSG00000090122	ENSMUSG00000019960	ENSMUSG00000039323
Monodelphis domestica – Opossum	ENSMODG00000017938		ENSMODG00000020516	ENSMODG00000015540
Birds				
Gallus gallus – Chicken	ENSGALG00000013078	ENSGALG00000023613	ENSGALG00000011207	ENSGALG00000011469
Meleagris gallopavo – Turkey	ENSMGAG00000013542		ENSMGAG00000011218	ENSMGAG00000011630
Taeniopygia guttata – Zebra finch				ENSTGUG00000005113
Reptiles				
Anolis carolinensis - Lizard	ENSACAG00000011754		ENSACAG00000011871	
Fish				
Danio rerio - Zebrafish	ENSDARG00000041104		ENSDARG00000070914	ENSDARG00000052470/ ENSDARG00000031422
Oryzias latipes - Medaka	ENSORLG00000009225			ENSORLG00000014695
Tetraodon nigroviridis - Pufferfish	ENSTNIG00000006917		ENSTNIG00000016051	ENSTNIG00000008865
Gasterosteus aculeatus – Stickelback				

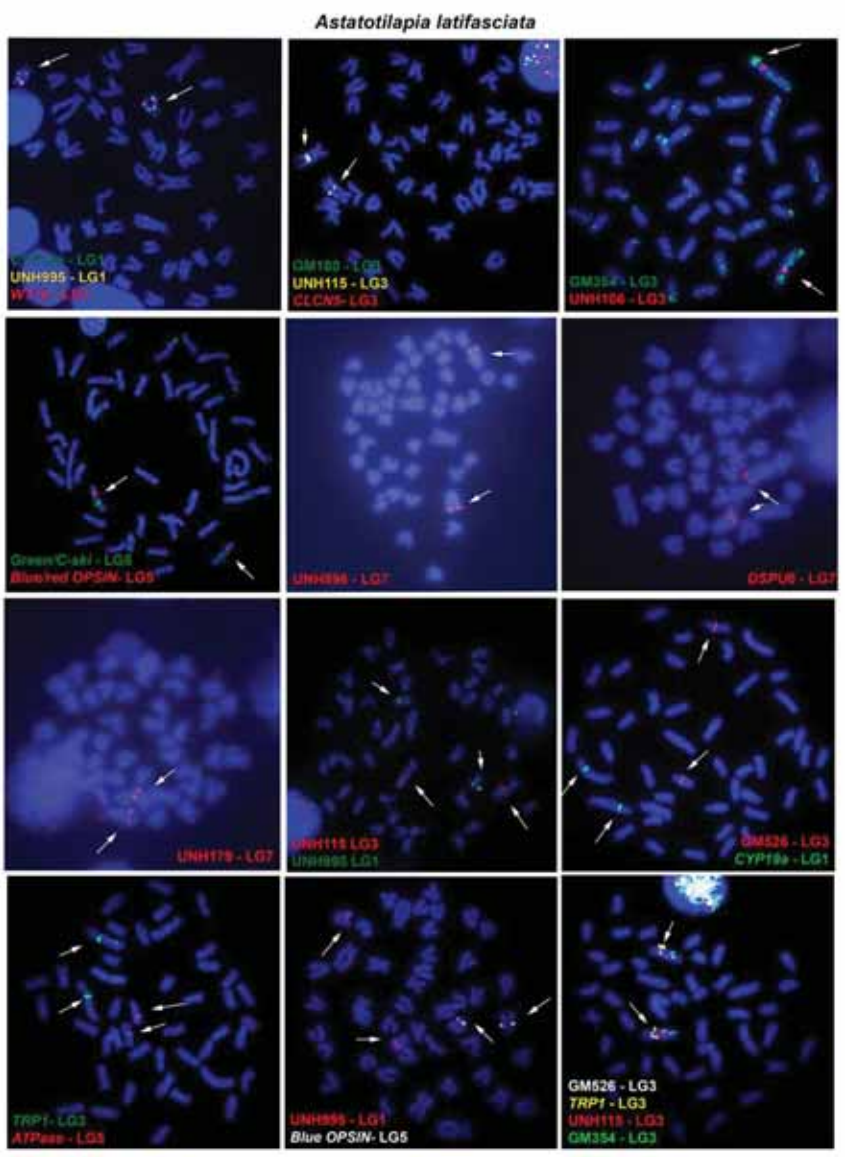
Supplementary material 1: Metaphase spreads of cichlid species probed with BAC clones from *Oreochromis niloticus*. BACs containing markers of LG 1, 3, 5 and 7 were hybridized under FISH procedure and are indicated in different colors in each metaphase. The arrows indicate the chromosome position of probes. Scale bar 10µm.

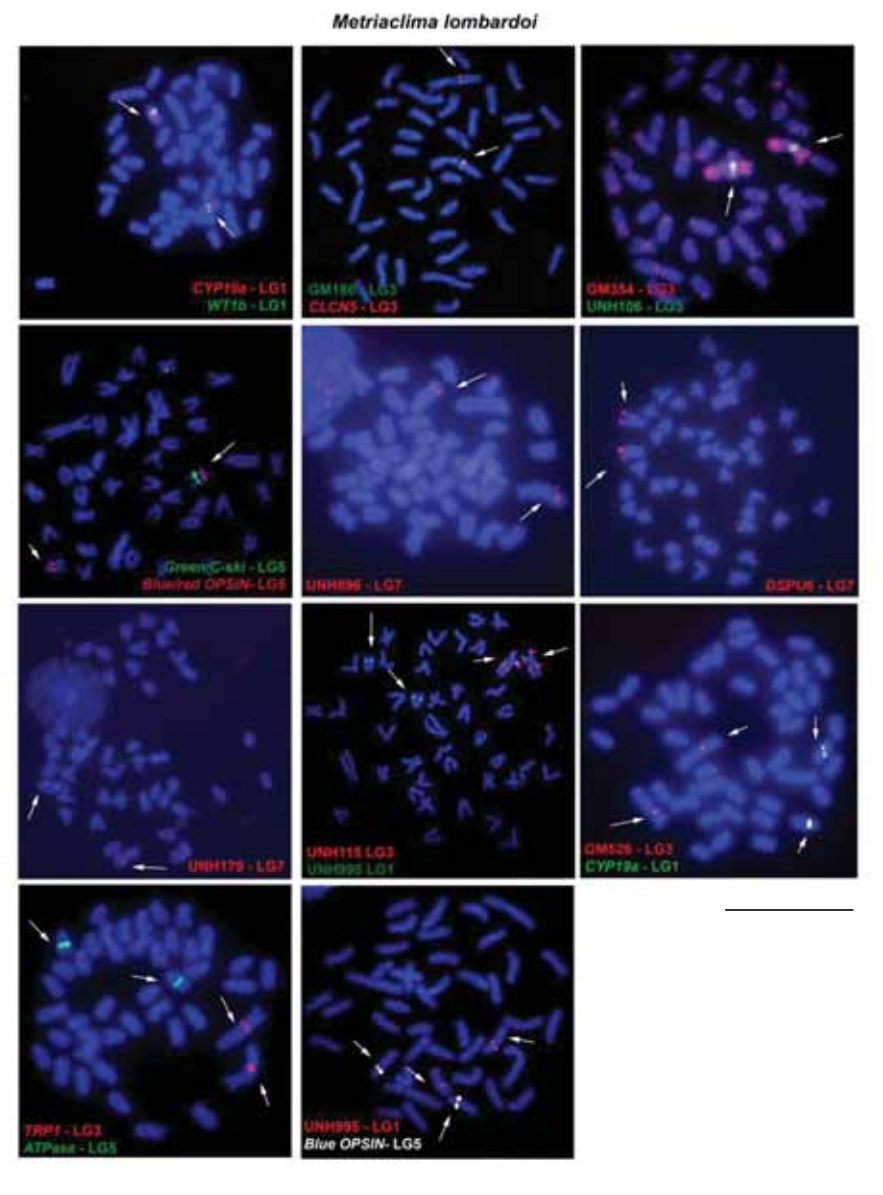


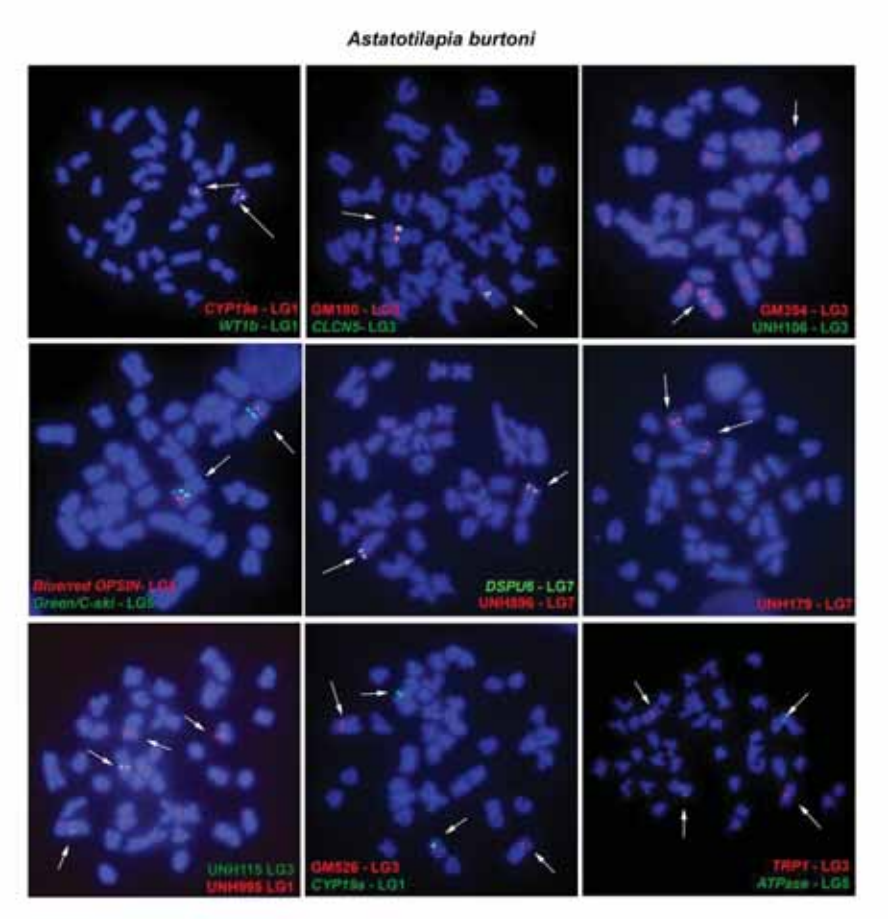

Oreochromis mossambicus
CYP19a - LG1
WT1b - LG1
GM180 - LG3
CLCN5 - LG3
GM354 - LG3
UNH106 - LG3
Bluetred OPSIN - LG5
Green/C-ski - LG5
UNH896 - LG7
DSPUR - LG7
UNH179 - LG7
UNH115 LG3
UNH995 LG1
GM526 - LG3
CYP19a - LG1
TRP1 - LG3
ATPase - LG5

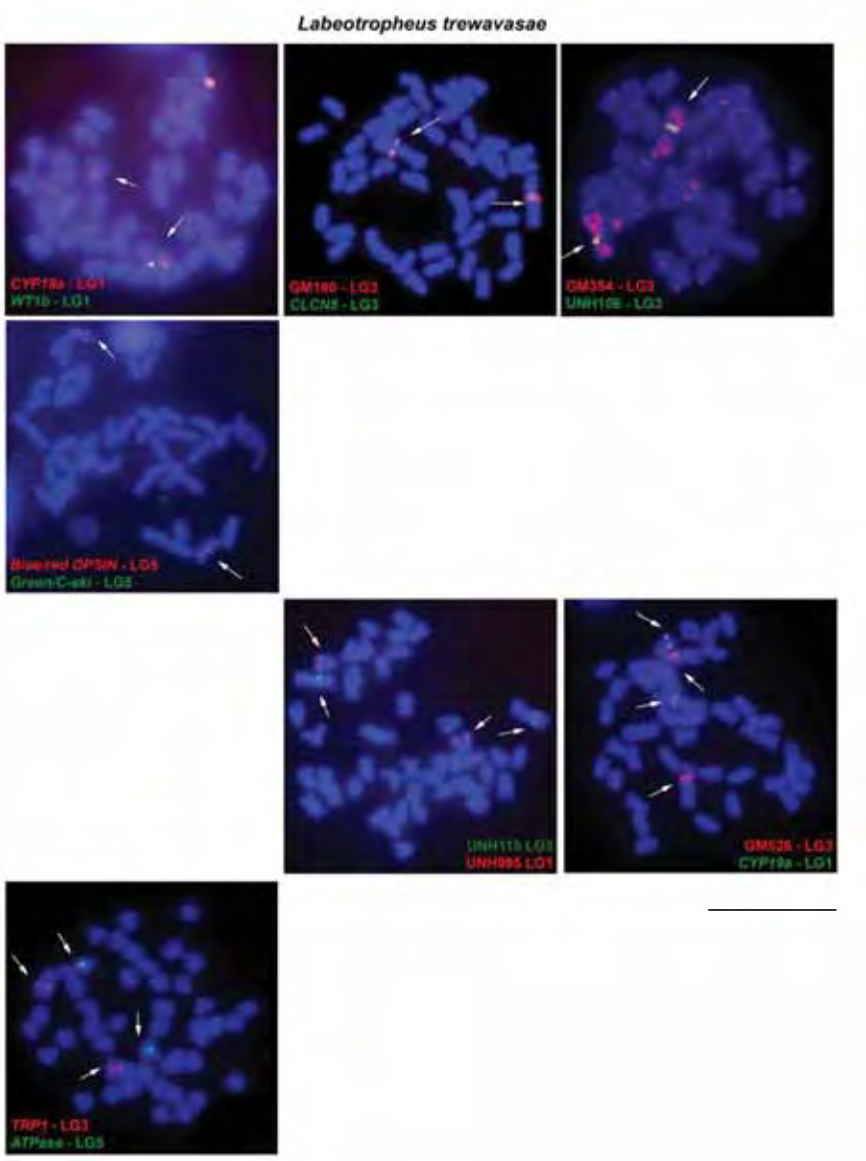

Oreochromis aureus
CYP19a - LG1
WT1b - LG1
GM180 - LG3
CLCN5- LG3
GM354 - LG3
UNH106 - LG3
Blue/red OPSIN- LG3
Green/C-sAI - LG5
DSPUII - LG7
UNH696 - LG7
UNH179 - LG7
UNH115 - LG3
UNH995 - LG1
GMI26 - LG3
CYP19a - LG1
TRPI - LG3
ATPase - LG5

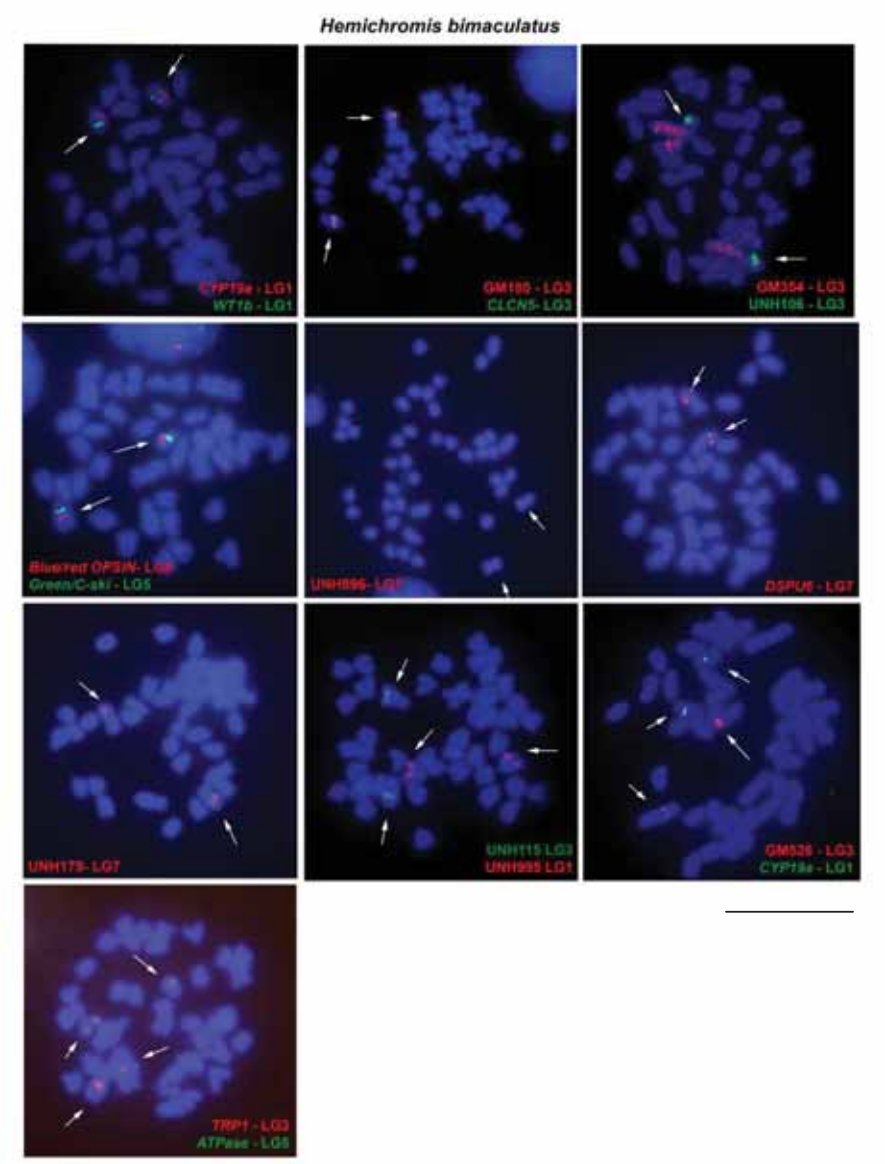

Tilapia mariae
CYP19a - LG1
-WT1B - LG1
GM180 - LG3
CLCN5- LG3
GM354 - LG3
UNH106 - LG3
Blue-cone OPSIN- LG5
Green/C-ski - LG5
DSPUI - LG7
UNH1179 - LG7
UNH115 LG3
GM526 - LG3
CYP19a - LG1
TRP1 - LG3
ATPase - LG5


Astatotilapia latifasciata
UNH995 - LG1
GM180 - LG3
UNH115 - LG3
CLCN5- LG3
GM354 - LG3
UNH106 - LG3
Green/C-ski - LG5
Blue/red OPSIN- LG5
UNH179 - LG7
UNH115 LG3
GM526 - LG3
CYP19a - LG1
TRP1 - LG3
ATPase - LG3
UNH995 - LG1
Blue OPSIN- LG5
GM526 - LG3
TRP1 - LG3
UNH115 - LG3
GM354 - LG3


Metriaclima lombardoi
CYP19a - LG1
WT1b - LG1
GM186 - LG3
CLCN5 - LG3
GM354 - LG3
UNH106 - LG3
Green/C-alt - LG5
Blue/red OPSIN- LG5
UNH896 - LG7
DSPU6 - LG7
UNH179 - LG7
UNH118 LG3
UNH895 LG1
GM529 - LG3
CYP19a - LG1
TRPI - LG3
ATPase - LG5
UNH995 - LG1
Blue OPSIN- LG5


Astatotilapia burtoni
CYP19a - LG1
WT1b - LG1
GM186 - LG3
CLCN5- LG3
GM334 - LG3
UNH106 - LG3
Green/C-akt - LG5
DSPU6 - LG7
UNH1179 - LG7
UNH115 LG3
GM526 - LG3
CYP19a - LG1
TRP1 - LG3
ATPase - LG5


Labeotropheus trewavasae
CYP19a - LG1
WT1b - LG1
GM186 - LG3
CLCN5 - LG3
GM354 - LG3
UNH108 - LG3
Blue-red OPSIN - LG5
Green/C-alt - LG5
UNH113 LG3
UNH995 LG1
GM526 - LG3
CYP19a - LG1
TRP1 - LG3
ATPase - LG5


Hemichromis bimaculatus
CYP19a - LG1
WT1b - LG1
GM185 - LG3
CLCN5- LG3
GM354 - LG3
UNH106 - LG3
Blue/red OPSIN- LG6
Green/C-ski - LG5
UNH896- LG7
DSPU6 - LG7
UNH179- LG7
UNH115 LG3
UNH995 LG1
GM526 - LG3
CYP19a - LG1
TRP1 - LG3
ATPase - LG5

4.2 Capítulo II:

Comparative cytogenetic mapping of *Sox*2 and *Sox14* in cichlid fishes and inferences on the genomic organization of both genes in vertebrates

Juliana Mazzuchelli[1], Fengtang Yang[2], Thomas D. Kocher[3] and Cesar Martins[1]

[1] Department of Morphology, Bioscience Institute, UNESP - São Paulo State University, 18618-000, Botucatu, SP, Brazil.

[2] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

[3] Department of Biology, University of Maryland, College Park, MD, 20742, USA.

Running title: Comparative mapping of *Sox* genes

Key words: Cichlidae, genome evolution, molecular cytogenetics, chromosome

Manuscrito aceito pela revista

Chromosome Research (2011) 19:657–667

DOI 10.1007/s10577-011-9225-4

Abstract

To better understand the genomic organization and evolution of *Sox* genes in vertebrates, we cytogenetically mapped *Sox2* and *Sox14* genes in cichlid fishes and performed comparative analyses of their orthologs in several vertebrate species. The genomic regions neighbouring *Sox2* and *Sox14* have been conserved during vertebrate diversification. Although cichlids seem to have undergone high rates of genomic rearrangements, *Sox2* and *Sox14* are linked in the same chromosome in the Etroplinae *Etroplus maculatus* that represents the sister group of all remaining cichlids. However, this genes are located on different chromosomes in several species of the sister group Pseudocrenilabrinae. Similarly the ancestral synteny of *Sox2* and *Sox14* has been maintained in several vertebrates, but this synteny has been broken independently in all major groups as a consequence of karyotype rearrangements that took place during the vertebrate evolution.

Introduction

Sox genes, a gene family encoding transcription factors involved in a variety of development processes, are found throughout the animal kingdom (Guo et al. 2009). These genes are expressed in various phases of embryonic development and, among vertebrates, are involved with testis development, neural crest cell development, neurogenesis, oligodendrocyte development, and chondrogenesis (Kiefer 2007). *Sox* genes are characterized by the presence of a DNA-binding HMG (high mobility group) domain and are subdivided into ten subgroups (A–J) based upon their HMG box sequences (Bowles et al. 2000). The *Sox* B group (*Sox1*, *Sox2*, *Sox3*, *Sox14* and *Sox21*) is of particular interest, since the members of this group play a major role in neural development and participate in the earliest events of central nervous system (CNS) differentiation in *Drosophila*, *Xenopus*, chicken and mouse (Collignon et al. 1996; Uchikawa et al. 1999; Hargrave et al. 2000; Kishi et al. 2000; McKimmie et al. 2005). Also they are most closely related to Sry and appear to be functionally conserved during evolution (McKimmie et al. 2005).

Sox1 and *Sox2* (group B1 members/activators), and *Sox14* and *Sox21* (B2 members/repressors), are arranged in two pairs, each comprising one *Sox*B1 activator and one *Sox*B2 repressor. At least one pair of group B1 and group B2 *Sox* genes are recognizable in the genomes of lower animals, including sponge (Larroux et al. 2008), sea urchin (Howard-Ashby et al. 2006), ascidian (Satou and Satoh 2005) and amphioxus (Meulemans and Bronner-Fraser 2007), although their linkages have not been confirmed. The early state of *Sox*B evolution is perfectly represented by the *Drosophila melanogaster* genome, which shows that the orthologs of vertebrates *Sox*2 and *Sox*14 are clustered on the same chromosome (Wei et al. 2011). This organization was also found in other insect genomes, including *Anopheles gambiae*, *Apis mellifera*, *Tribolium castaneum*, *Nasonia vitripennis* and

Bombyx mori, where *Sox* B genes are clustered on the same chromosome or scaffold assembly (Wei et al. 2011). Among vertebrates, *Sox* B genes have arisen by rounds of whole genomic duplications, rearrangement and divergence from ancestral *Sox* B genes (Kirby et al. 2002; Guth and Wegner 2008).

Linkage of *Sox2* and *Sox14* was previously observed in several mammals and in the chicken *Gallus gallus* (revised in Popovic and Stevanovic 2009). At the same time, *Sox2* and *Sox14* were not linked in other mammals and in the fish *Danio rerio* (revised in Popovic and Stevanovic 2009). In the cichlid fish *O. niloticus*, *Sox2* and *Sox14* genes mapped to different chromosomes corresponding to linkage group 17 (LG17) and LG23, respectively (Cnaani et al. 2007). Although both conditions of linkage and non linkage of *Sox2* and *Sox14* seem to occur in different mammals, the absence of information for non-mammalian species does not allows major conclusions. Okuda et al. (2006) suggest that the chromosomal organization of group B *Sox* genes in fishes is different from other vertebrates. It is not yet clear whether the duplication, coupled with functional divergence and physical dispersal occurred early in the radiation of vertebrates, or occurred more recently.

Cichlids have been used as model organisms to study a diversity of evolutionary mechanisms because they represent one of the most striking examples of rapid and convergent evolutionary radiation among vertebrates. Here, we cytogenetically mapped the distribution of *Sox2* and *Sox14* on the chromosomes of several cichlid species by fluorescence *in situ* hybridization (FISH) and performed a comparative mapping analyses of *Sox2* and *Sox14* orthologs in numerous vertebrate species based on the available genomic databases. Our results show that the linkage of *Sox2* and *Sox14* is maintained in many vertebrate taxa and that the separation of these genes onto different chromosomes seems to have occurred independently in all major vertebrate groups.

Materials and methods

Animals and sampling

Cichlid species used in this work were obtained from three sources. Cichlids from Lake Malawi were collected from the wild from 2005-2008 and maintained in the Tropical Aquaculture Facility (TAF) of the University of Maryland (UMD), College Park, MD, USA. South American species were collected from the wild in several Brazilian rivers. Additional species of uncertain origin were obtained from commercial sources in Botucatu, SP, Brazil, and were maintained in the Fish Room of the Laboratório Genômica Integrativa (FR-LGI) at São Paulo State University (UNESP), Botucatu (Table 1). All the specimens examined were fixed in formaldehyde and then stored in alcohol in the fish collections of TAF-UMD and FR-LGI.

Bacterial artificial chromosome (BAC) clones and probe labelling

Two BAC clones, from a genomic library of the Nile tilapia *O. niloticus*, containing the genes *Sox2* from LG23 and *Sox14* from LG17 (BAC IDs b04TI053B06 and b03TI079I04, respectively) (Cnaani et al. 2007) were used as probes for fluorescence *in situ* hybridization (FISH). BAC extraction was conducted using the PhasePrep®^{TM} BAC DNA Kit (Sigma-Aldrich, St Louis, MO, USA) according to supplier's protocol. The BAC clones were labeled with biotin or digoxigenin (DIG) coupled nucleotides (Roche Applied Sciences, Indianapolis, IN, USA) using whole genome amplification (WGA2 &3) kits (Sigma-Aldrich), according to the supplier's protocol. For double-color FISH, 16 µl of a hybridization mixture containing 50% deionized formamide, 2XSSC, 10% dextran sulfate, 10 µg of salmon sperm DNA, and 100 ng each of biotin and dig-labeled probes was prepared, denatured for 10 min at 65°C and immediately cooled on ice.

Chromosome preparation and FISH procedure

Chromosome preparations were obtained as previously described (Bertollo et al. 1978), and the slides with the chromosomes were air-dried, treated with pepsin (0.01% in 10 mM HCl) and dehydrated in an ethanol series one day before use. The slides were denatured in 70% formamide/2xSSC, pH 7 for 40s, and dehydrated in an ice-cold ethanol series. 16µl of probe mixture (containing 100 ng of each DNA probe) were hybridized under a 22 mm X 32 mm cover slip in a 37°C moist chamber for 48 h. Slides were washed two times for 5 min in 50% formamide/2xSSC, pH 7 at 43°C with agitation, then 10 min in 2xSSC, pH 7 at 42°C with continuous agitation. Hybridization signals were detected with avidin-FITC and rhodamine-anti-DIG (Roche Applied Sciences, Indianapolis, IN, USA), according to the supplier's protocol. After three washes of 2 min in phosphate buffer detergent (4xSSC/1% Tween-20), slides were mounted with antifade solution containing DAPI. Results were recorded with an Olympus BX61 microscope equipped with an Olympus digital camera DP71 and software Image-Pro MC 6.0.

Sequences similarity and comparative genomic database analyses

The sequences of an 883 nucleotide fragment of *Sox*2 and the entire coding sequence of *Sox14* from *Oreochromis niloticus* were obtained from National Center for Biotechnology Information (NCBI) databases (www.ncbi.nlm.nih.gov/genbank/) with accession numbers EF431920–EF431927) (Cnaani et al. 2007). The similarity analyses between several vertebrates were done using nucleotide megablast at Basic Local Alignment Search Tool (BLAST) at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The chromosomal locations of *Sox*2 and *Sox*14 genes among several vertebrates were determined using information currently available in the public genomic databases of NCBI

Map Viewer (http: //www.ncbi.nlm.nih.gov/genomes), Sanger Institute Ensembl Database (http://www.ensembl.org) and BouillaBase-Comparative Genome Browsers (www.BouillaBase.org). The syntenic relationship analyzes of *Sox*2 and *Sox*14 genes were conducted using Genomicus genome browser (http://www.dyogen.ens.fr/genomicus/). Each gene was analyzed separated using human and zebrafish as reference species. The identification of syntenic genes among fish species using *O. niloticus* as a reference was determined using BouillaBase browser, since the gene prediction of *O. niloticus* is not yet available in the Genomicus browser.

Results

Comparative cytogenetic mapping

The *Sox2* and *Sox14* genes mapped to different chromosomes in *O. niloticus* (Figure 1), corresponding to LG17 and LG23 respectively, as expected (Cnaani et al. 2007). In the other eight Pseudocrenilabrinae species studied, *Sox2* and *Sox14* also mapped on two different chromosome pairs (Figure 1). The *Sox*2 (LG17) was located in the pericentromeric region of a medium subtelocentric/acrocentric (st/a) chromosome in all African cichlids investigated (Figure 1). *Sox*14 (LG23) showed variations in the chromosomal position being interstitially located on the long arm of a larger st/a chromosome in all tilapiine species (Figure 1) and in the haplochromine species (Figure 1). Exceptions were observed in *Labeotropheus trewavasae* (haplochromine), where *Sox14* was located on a meta/submetacentric (m/sm) chromosome (Figure 1) as well as in the hemichromine, *Hemichromis bimaculatus* (Figure 1). B chromosomes were detected in *Haplochromis obliquidens* and *Metriaclima lombardoi* (haplochromines) as previously reported (Poletto et

al. 2010a; 2010b), but no signal of *Sox* genes were detected in these B chromosomes (data not shown).

In the Asian cichlid species, *Etroplus maculatus* (Etroplinae), the *Sox2* and *Sox14* were positioned on the two arms of a single small metacentric chromosome pair (Figure 2). In South American cichlids (Cichlinae) belonging to different tribes (Table 1), none of the BAC probes produced identifiable chromosomal signals, probably because the occurrence of rearrangements that could have differentiated the genomic blocks containing *Sox2* and *Sox14* in relation to the *O. niloticus*, the species source of the BAC clones used as the chromosome probes.

Sox genes of cichlids and comparative genomics

Analysis of similarity between *Sox2* (EF431924.1) and *Sox14* (EF431920.1) gene sequences of *Oreochromis niloticus* (Cnaani et al. 2007) indicates high levels of conservation among the homologs *Sox* genes of other vertebrate species, ranging from 79% to 100% for *Sox*2 (See Supplementary Material S1) and 78% to 100% for *Sox14* (See Supplementary Material S2). Furthermore, based on web databases we have determined that *Sox2-Sox14* are linked in several vertebrate species, including some mammals (*Gorilla gorilla*, *Pongo abelli*, *Sus scrofa*, *Bos taurus* and *Ornithorhynchus anatinus*) and the bird *Gallus gallus* (Table 2). On the other hand, *Sox2* and *Sox14* are located on different chromosomes in other vertebrates, including mammals (*Callithrix jacchus, Canis familiaris, Equus cabalus*, *Rattus norvegicus, Mus musculus*, *Monodelphis domestica*) and the fish *Danio rerio* (Table 2). The genomic position of *Sox2* was also identified for the fish species *Tetraodon nigroviridis*, *Gasterosteus aculeatus* and *Oryzias latipes*, but no information was retrieved for *Sox14* (Table 2), probably because more genome sequence and physical chromosome map data are available related to

Sox2 than to *Sox14*. For several other species, it was impossible to determine the genomic organization of both *Sox2* and *Sox14* genes because the existing genome sequences are incomplete.

Analyses of the syntenic relationship between vertebrate chromosomal segments containing *Sox2* and *Sox14* genes were conducted using the Genomicus genome browser. The analyses of each gene separately using *Homo sapiens* as the reference species show that synteny is highly conserved in the chromosomal segments containing the *Sox* genes. A large genomic block containing several genes was conserved through vertebrates (Figure 3). At least 4 genes around *Sox*2 region are conserved even between more distant species like mammals and fishes: TTC14, FXR1, DNAJC19 and MCCC1 genes are present at least in three fish species and human (Figure 3). The *Sox14* region seems to be more divergent between mammals and fishes, and sometimes the gene correspondence was not clear (see medaka and *Tetraodon*, for example, in Figure 3) or even the synteny correspond to a unique gene (gene CLDN18 in fugu and gene DZIP1L in zebrafish for example, Figure 3). Considering that *O. niloticus* gene content is not yet available in the Genomicus browser, the genomic block containing *Sox2* was comparatively analyzed among several fish species using BouillaBase (Figure 4). In this new analysis several genes were detected conserved between *O. niloticus*, stickleback, medaka, *Takifugu* and *Tetraodon* (Figure 4). On the other hand, zebrafish presented few conserved genes with the other fish species. Furthermore, two genes (FXR1 and TTC14) were observed in most fish species (Figure 4) and also in several non-fish vertebrates (Figure 3). That analysis was not possible for *Sox14* because the genome annotation for *O. niloticus* is not yet complete.

Discussion

General aspects on the genome organization of Sox2 and Sox14 genes

The analysis of syntenic regions of the chromosomal locations that harbour the *Sox* genes using Genomicus have demonstrated that some genes which flank the *Sox2* and *Sox14* orthologs are conserved in their positions in some mammalian species. *Sox14* is more conserved (only few rearrangements were detected) than *Sox* 2 (more rearrangements were observed) among mammals. However, when the *Sox14* regions were compared using diverse groups (mammals, birds, fishes), they were not conserved as observed in mammals. *Sox14* orthologs are highly diverged in non-mammal groups and *Sox2* orthologs are more stable among all vertebrates.

The analysis of the genomic blocks containing *Sox* genes suggest the genes observed in the region are evolving as part of a large block of genes rather than individually. This is clearly observed among mammals but not much clear for fishes maybe because the (i) limited amount of genomic data available or (ii) the intense dynamism that rules the genome evolution in teleost fishes. The size of the syntenic blocks looks smaller in fishes, maybe because the low level of sequence similarities in distant comparisons has made it difficult to identify unambiguously orthologs, or the loss of *Sox*2 and *Sox14* regions. Even using only fish species in the comparative analysis, the size of the syntenic blocks are still limited, but it is possible to detect some genes still present in mammals such as FXR1 and TTC14 (see Figures 3 and 4).

In the pufferfish *Takifugu rubripes*, 25 *Sox* genes were identified occurring as duplicated paralogs with the mammalian *Sox*1, *Sox*4, *Sox*6, *Sox*8, *Sox*9, *Sox*10 and *Sox*14 (Koopman et al. 2004). In contrast, *Sox*8 and *Sox*10 are not duplicated in the zebrafish. However, zebrafish has three *Sox*B2 genes (*Sox14*, *Sox21a* and *Sox21b*) and six *Sox*B1 genes

(*Sox1a*, *Sox1b*, *Sox*2, *Sox3*, *Sox19a* and *Sox19b*) (Okuda et al. 2006). The second copies of *Sox*2 and *Sox*3 might have been lost early in the teleost lineage, because the *Takifugu* genome also contains only one copy of *Sox*2 and *Sox3* (Koopman et al. 2004). The same should be occurring with cichlids, where the second copies of *Sox*2 and *Sox14* are absent and may have been lost very early in the teleost fish radiation. Interestingly, there is no direct *Sox*B ortholog for teleost *Sox*19a/b in other vertebrates, so these are fish specific genes. Instead, the highly divergent mammalian *Sox*G gene *Sox15* and *Xenopus Sox*D appear to be the closest relatives to *Sox*19a/b (Okuda et al. 2006).

Cytogenetic mapping of Sox2 and Sox14 in cichlids and inferences on the chromosomal rearrangements involving both genes

Etroplus maculatus contains the *Sox2* and *Sox14* genes preserved in the same chromosome (Figure 2 and 5), i.e. the ancestral vertebrate condition previously suggested (Kirby et al. 2002). The subfamily Etroplinae is considered the sister group of all remaining cichlids (Smith et al. 2008) and we could propose that these genes were syntenic in the ancestor of cichlids and were preserved linked on the same chromosome of Etroplinae species until now. On the other hand, cytogenetic mapping of *Sox2* and *Sox14* in Psedocrenilabrinae cichlids suggests that chromosomal rearrangements during the diversification of this group separated *Sox2* and *Sox14* genes onto different chromosomes (Figure 1 and 5). Although a higher number of species should be analyzed, the data obtained for the species here investigated suggest the chromosome that harbours *Sox2* has been conserved during Pseudocrenilabrinae diversification. However, variations in the morphology of the chromosome carrying *Sox14* among Pseudocrenilabrinae species were observed,

suggesting that this chromosome has undergone more rearrangements during the evolution of the group (Figure 5).

The chromosomal organization of these two genes among several vertebrates apparently does not follow a unique pattern. Human *Sox2* and *Sox14* are linked on chromosome 3 and map together in the platypus as well (Hope et al. 1990), demonstrating the synteny conservation of *Sox*2 and *Sox14* over at least 170 million years since mammalian groups Prototheria and Theria diverged. *Sox2-Sox14* maps together also in several other primates like, *Gorilla gorilla*, *Macaca mulatta, Pan troglodytes* and *P. abelli,* in *Sus scrofa* (pig), in *Oryctolagus cuniculus* (rabbit), and in *Bos taurus* (cattle) (Popovic and Stevanovic 2009, present work). On the other hand, no linkage was found for *Sox2-Sox14* pair in dog, mouse, and in the primate *Callitrix jacchus* (Popovic and Stevanovic 2009, present work).

The present analysis suggests that the expected ancestral linkage for *Sox2* and *Sox14* is maintained in diverse vertebrate taxa and the genomic split of these genes to different chromosomes occurred independently in all major vertebrate groups (Figure 6) may be a consequence of particular karyotype rearrangements such as translocation or transposition, for example. The presence of linkage and non-linkage of *Sox2* and *Sox14* within Cichlidae suggests that the events of separation of both genes also occurs in the terminal taxa level (recent evolutionary events) and are not only restrict to major vertebrate groups (ancient evolutionary events).

The integration of cytogenetic mapping and comparative genomics of *Sox* genes in other vertebrates would further improve our understanding of the structure, organization and evolution of *Sox* genes. Unfortunately deep analysis integrating cytogenetics and genomic data were not possible for cichlids because there is no large scale genomic data available for the family yet. Although small chromosome variations were observed in the location of *Sox2*

among Pseudocrenilabrinae cichlids, it seems that *Sox14* occupies a more dynamic genomic region resulting in variations in its chromosomal position among the species.

Acknowledgments

This study was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) from Brazil.

Tables

Table 1: Cichlids analyzed.

Subfamily and wild distribution	**Groups or Tribes**	**Species**	**2n**	**Origin of specimens**
Etroplinae (India and Madagascar)		*Etroplus maculatus*	46	Petshop
Pseudocrenilabrinae (Africa)	Tilapiine	*Oreochromis niloticus*	44	TAF-UMD
		Oreochromis mossambicus	44	TAF-UMD
		Oreochromis aureus	44	TAF-UMD
		Tilapia mariae	40	TAF-UMD
	Haplochromine	*Haplochromis obliquidens*	44	Petshop
		Metriaclima lombardoi	44	TAF-UMD
		Astatotilapia burtoni	40	TAF-UMD
		Labeotropheus trewavasae	44	TAF-UMD
	Hemichromine	*Hemichromis bimaculatus*	44	Petshop
Cichlinae (America)	Cichlini	*Cichla kelberi*	48	Araguaia River, Brazil
	Astronotini	*Astronotus ocellatus*	48	Tiête River, Brazil
	Heroini	*Symphysodon aequefasciatus*	48	Petshop
	Geophagini	*Geophagus brasiliensis*	48	Tietê River, Brazil

Table 2: Chromosomal position of *Sox2* and *Sox14* genes in different vertebrate species and their nucleotide similarity level compared to *O. niloticus*. Chr, chromosome position; S *On*, Similarity to *O. niloticus*; AN, Accession number; NA, sequences for *Sox* genes are not available; NM, sequences of *Sox* genes are available but it was not possible to identify the genomic position. Species with linkage of *Sox2* and *Sox14* are highlighted in yellow, species in which *Sox2* and *Sox14* are unlinked are highlighted in blue, and the species with the linkage data not yet determined for *Sox2* and *Sox14* are highlighted in red. *Species whose *Sox2* and *Sox14* linkage was previously checked/revised in Popovic and Stevanovic 2009.

Major group/Species	***Sox*2**			***Sox*14**		
	Chr	**S *On***	**AN**	**Chr**	**S *On***	**AN**
Mammals						
Homo sapiens *	3	81%	NG_009080	3	79%	NM_004189
Pan troglodytes *	3	83%	XM_516895	3	79%	XM_526317
Gorilla gorilla	3		ENSFM00500000270951	3		ENSGGOG00000013878
Pongo abelii	3	82%	XM_002814321	3	79%	XM_002814084
Callithrix jacchus	15	82%	XM_002807565	1	83%	XM_002742479
Macaca mulatta *	2	81%	NM_001142940	2	80%	NM_001194657
*Canis familiaris**	34	82%	XM_545216	23		ENSCAFG00000009867
Bos taurus *	1	81%	NM_001105463	1	78%	NM_001163781
Equus cabalus *	19	80%	NM_001143799	16	80%	XM_001916428
Ornithorhynchus anatinus *	1	81%	XM_001506934	1		AY112710
Sus scrofa	13	82%	EU503117	13		ENSSSCG00000011656
Oryctolagus cuniculus	14	82%	XM_002716451	14		ENSOCUG00000001686
Loxodonta africana	NM		ENSLAFG00000006362	NM		ENSLAFG00000003447
Rattus norvegicus *	2	81%	NM_001109181	8		NW047801.1
Mus musculus *	3	81%	NM_011443	9	78%	NM_011440
Monodelphis domestica *	7	85%	XM_001368783	4		ENSMODG00000024983
Cavia porcellus	NM		ENSCPOG00000003575	NM		ENSCPOG00000026272
Echinops telfairi	NM		ENSETEG00000005122	NM		ENSETEG00000018576
Birds						
Gallus gallus *	9	79%	D50603	9		ENSGALG00000017372
Amphibians						
Xenophus tropicalis	NM	86%	BC159121	NM		ENSXETG00000022689
Fish						
Danio rerio *	22	82%	AB242329	6		ENSDARG00000070929
Oryzias latipes	4	93%	FJ895588	NM	91%	NM_001164872
Tetraodon nigroviridis	1		ENSTNIG00000008596	NM	100%	AY612092
Gasterosteus aculeatus	8		ENSGACG00000020111	NA		NA

Figure captions

Fig. 1 Cytogenetic mapping of *Sox2* and *Sox14* (arrows) in Pseudocrenilabrinae Cichlidae species showing their distribution in different chromosomes. The tilapiines include *Oreochromis niloticus, Oreochromis mossambicus*, *Oreochromis aureus* and *Tilapia mariae*; the haplocharomine includes *Haplochromis obliquidens, Metriaclima lombardoi, Astatotilapia burtoni* and *Labeotropheus trewavasae*; and the hemichromines is represented by *Hemichromis bimaculatus*. Scale bar 5μm.

Fig. 2 Cytogenetic mapping of *Sox2* (green label) and *Sox14* (red label) genes (arrows) in *Etroplus maculatus*. In the insert, a chromosome labeled pair of a second metaphase spread. Both genes are positioned on the same chromosome. Scale bar 5μm.

Fig. 3 Chromosomal segments showing the conserved syntenic blocks containing *Sox2* and *Sox14* genes in diverse vertebrates and only in fish. Color squares indicate the same gene in the different vertebrate species (left) and its respective genomic position in relation to several other genes (right). The most conserved genes among fish and other vertebrates are underlined. See Supplementary Material S3 for more information on the gene abbreviations.

Fig. 4 Chromosomal segments showing the conserved syntenic blocks containing *Sox2* in several fish species, including the cichlid *O. niloticus*. Vertical color bars indicate the same gene in the different species (left) and its respective genomic position in relation to several other genes (right). The most conserved genes among fish and other vertebrates are underlined. The interrupted vertical lines indicate the genomic positions of the genes in

reference to the scaffold 243 of *O. niloticus* genome that was used as reference. See Supplementary Material S4 for more information on the gene abbreviations.

Fig. 5 Phylogenetic relationship of cichlids (adapted from Smith et al. 2008) showing the chromosomal distribution of *Sox2* and *Sox14* genes.

Fig. 6 Evolutionary perspective for the distribution of *Sox2* and *Sox14* among vertebrates. Rounds of whole genomic duplications, rearrangement and divergence from ancestral *Sox* B genes have originated the actual scenario observed.

Figure 1

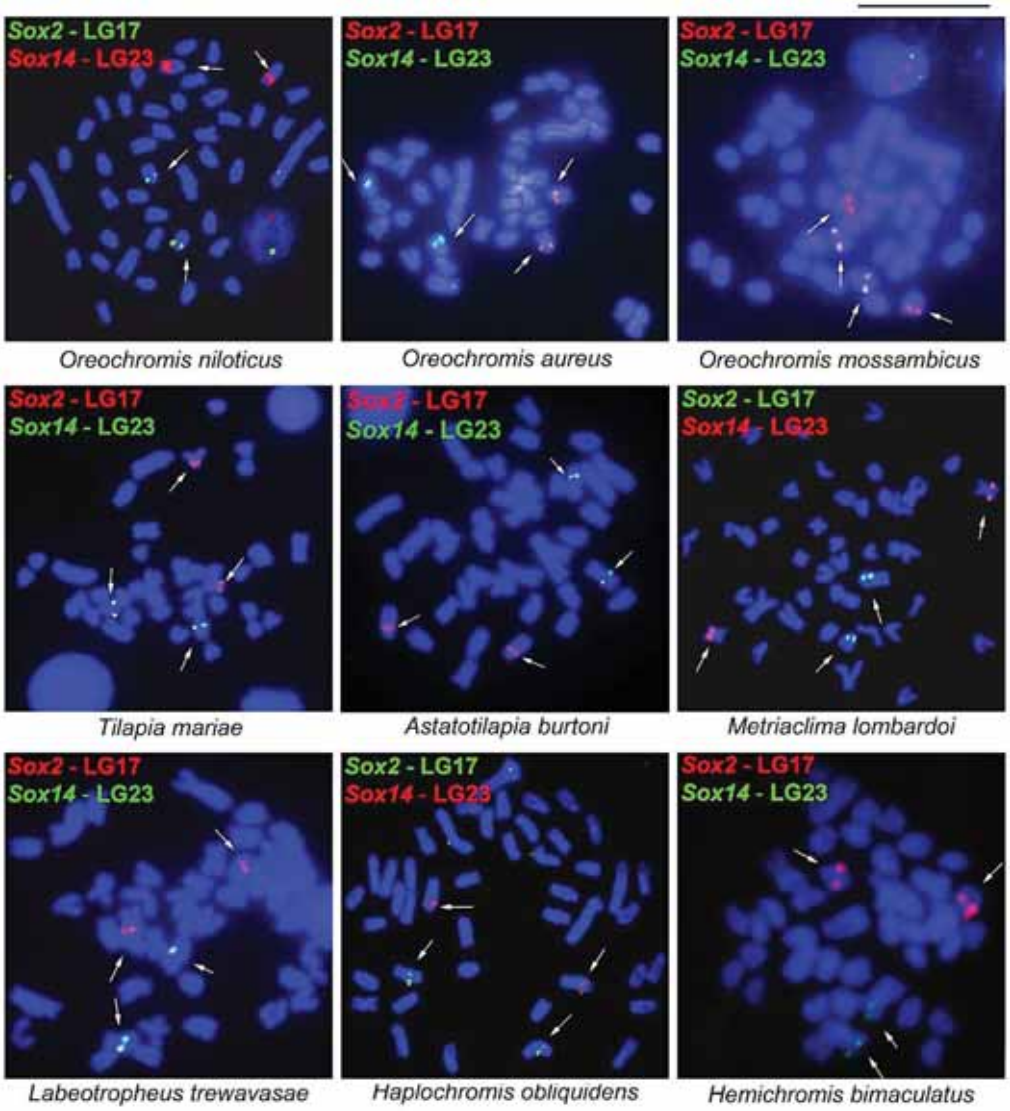


Figure 2

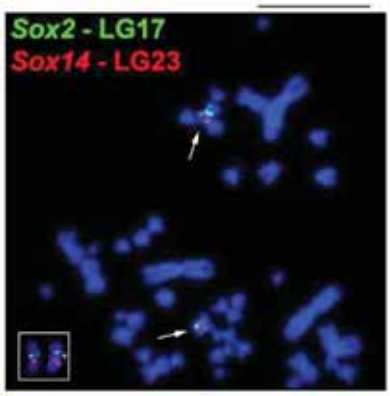


Figure 3

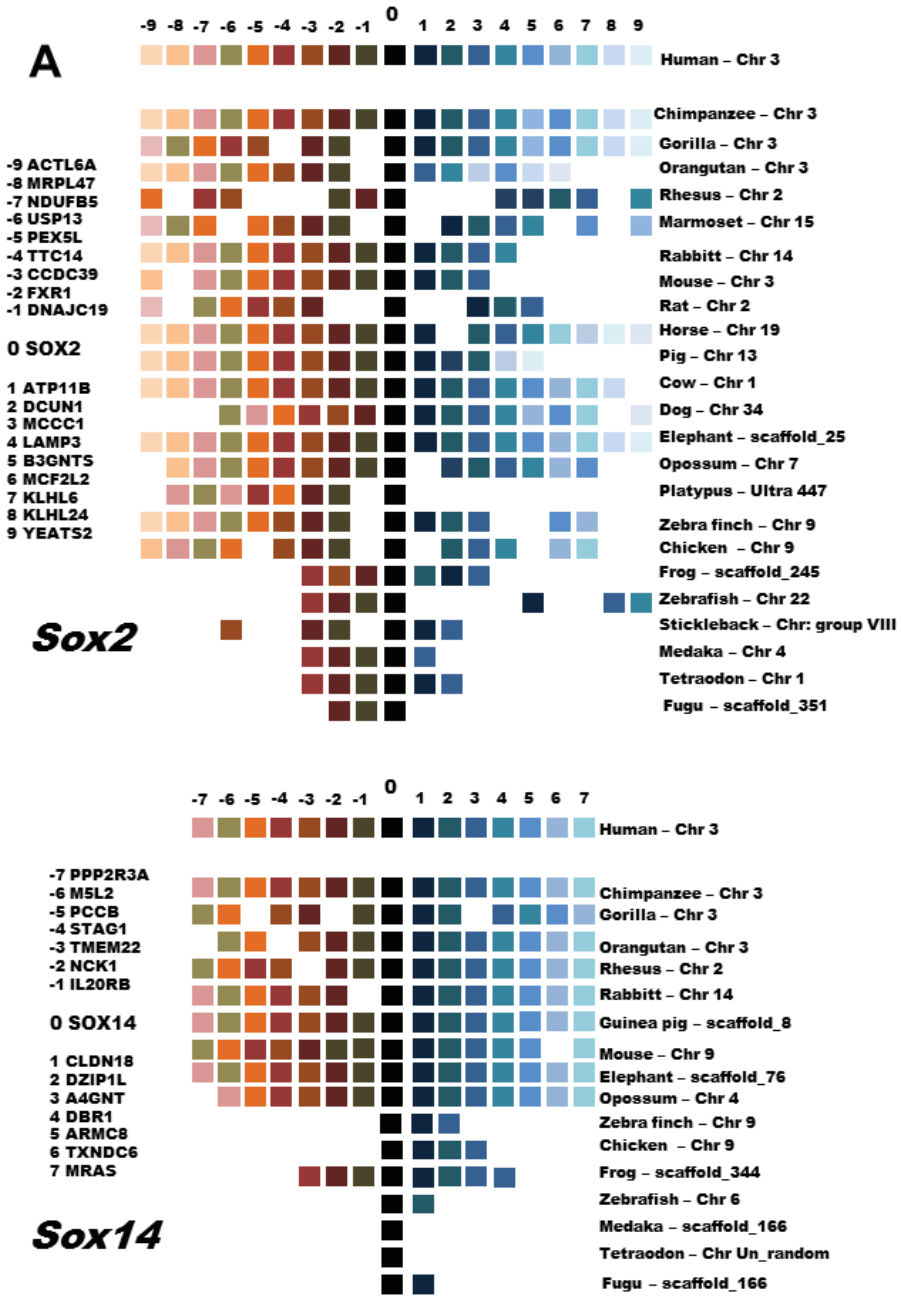

A
-9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9
-9 ACTL6A
-8 MRPL47
-7 NDUFB5
-6 USP13
-5 PEX5L
-4 TTC14
-3 CCDC39
-2 FXR1
-1 DNAJC19
0 SOX2
1 ATP11B
2 DCUN1
3 MCCC1
4 LAMP3
5 B3GNTS
6 MCF2L2
7 KLHL6
8 KLHL24
9 YEATS2
Sox2
Human – Chr 3
Chimpanzee – Chr 3
Gorilla – Chr 3
Orangutan – Chr 3
Rhesus – Chr 2
Marmoset – Chr 15
Rabbitt – Chr 14
Mouse – Chr 3
Rat – Chr 2
Horse – Chr 19
Pig – Chr 13
Cow – Chr 1
Dog – Chr 34
Elephant – scaffold_25
Opossum – Chr 7
Platypus – Ultra 447
Zebra finch – Chr 9
Chicken – Chr 9
Frog – scaffold_245
Zebrafish – Chr 22
Stickleback – Chr: group VIII
Medaka – Chr 4
Tetraodon – Chr 1
Fugu – scaffold_351
-7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7
-7 PPP2R3A
-6 M5L2
-5 PCCB
-4 STAG1
-3 TMEM22
-2 NCK1
-1 IL20RB
0 SOX14
1 CLDN18
2 DZIP1L
3 A4GNT
4 DBR1
5 ARMC8
6 TXNDC6
7 MRAS
Sox14
Human – Chr 3
Chimpanzee – Chr 3
Gorilla – Chr 3
Orangutan – Chr 3
Rhesus – Chr 2
Rabbitt – Chr 14
Guinea pig – scaffold_8
Mouse – Chr 9
Elephant – scaffold_76
Opossum – Chr 4
Zebra finch – Chr 9
Chicken – Chr 9
Frog – scaffold_344
Zebrafish – Chr 6
Medaka – scaffold_166
Tetraodon – Chr Un_random
Fugu – scaffold_166

Figura 4

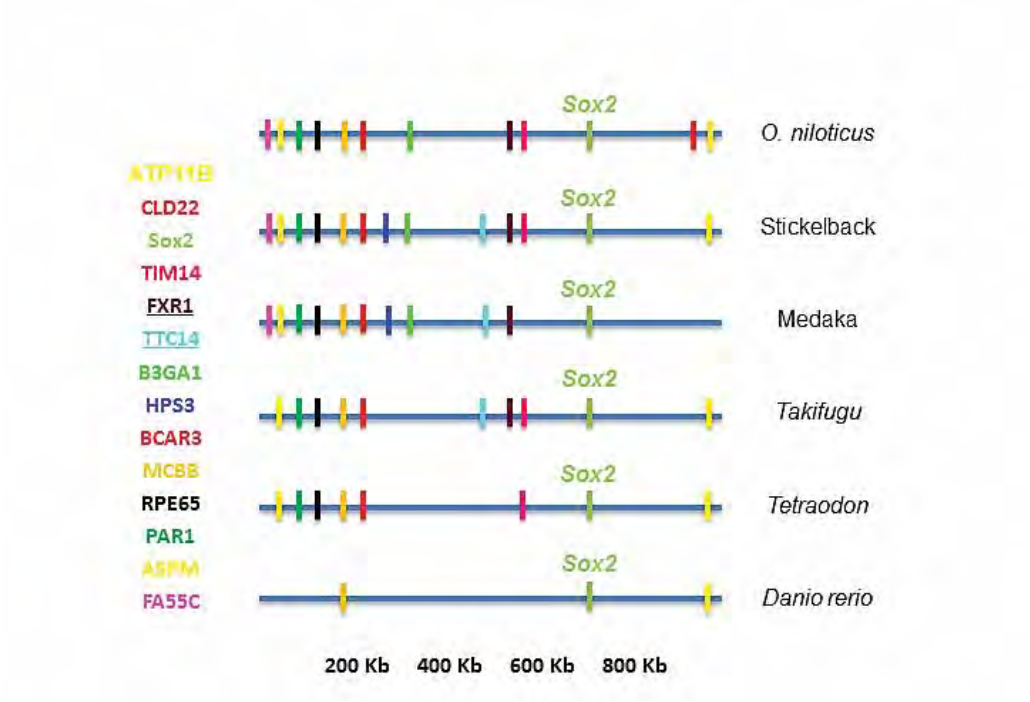


Figure 5

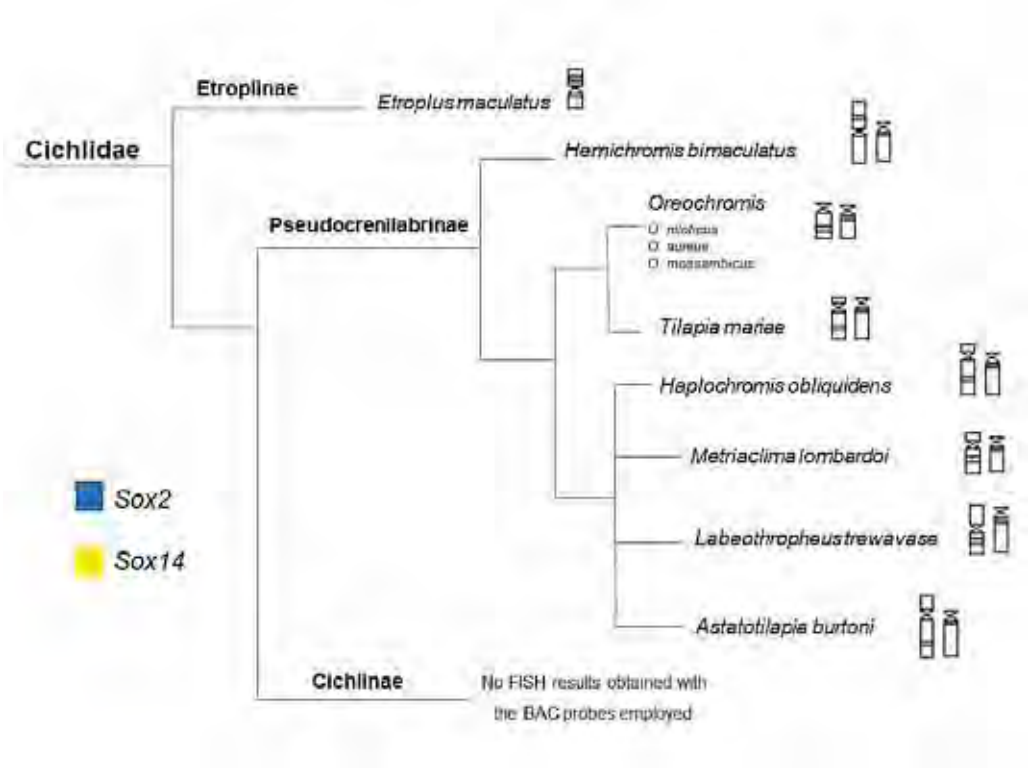
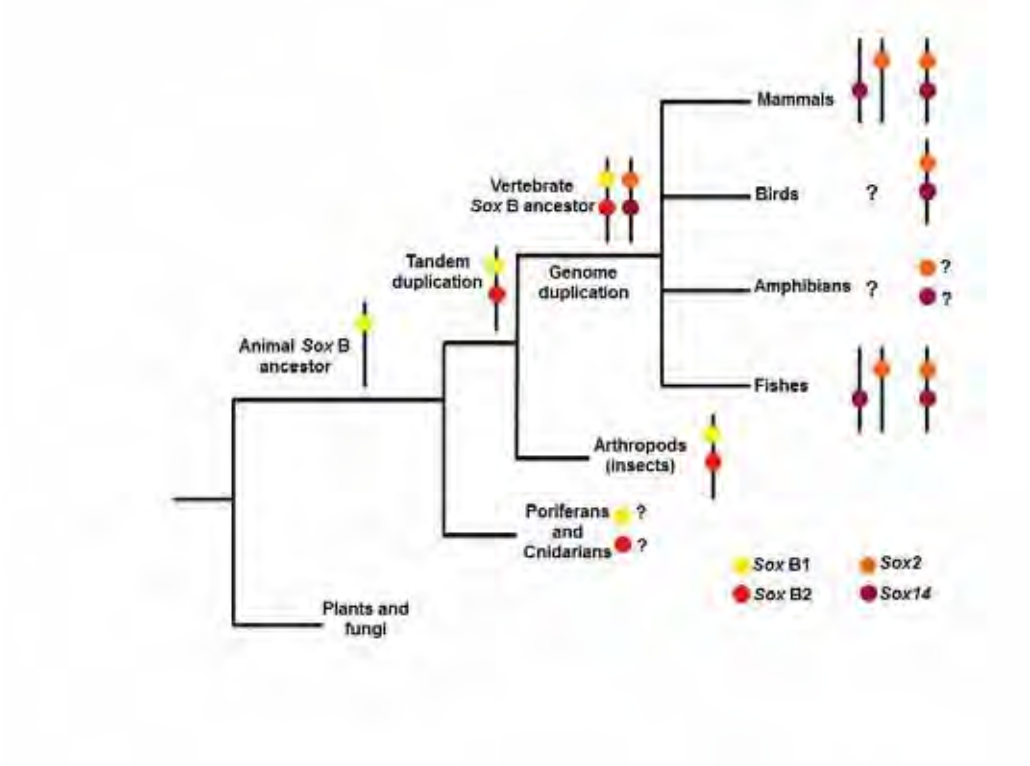


Figure 6


Mammals
Birds
?
Vertebrate
Sox B ancestor
Tandem
duplication
Genome
duplication
Amphibians
?
?
?
Animal Sox B
ancestor
Fishes
Arthropods
(insects)
Poriferans
and
Cnidarians
?
?
Sox B1
Sox B2
Sox2
Sox14
Plants and
fungi

Supplementary material 1: Similarity of *Sox2* gene between *O. niloticus* and several vertebrates based on Blast analysis at NCBI.

Accession number	**Sequence information**	**Match score**
EF431921.1	*Oreochromis aureus Sox*2 (*Sox*2) gene, partial cds	100%
FJ432695.1	*Epinephelus coioides* cell-line GBC4 *Sox*2 mRNA, partial cds	94%
GU001787.1	*Trachinotus blochii Sox*2 mRNA, partial cds	94%
FJ895588.1	*Oryzias latipes* SRY-box containing gene 2 (*Sox*2) mRNA, partial cds	93%
AY277952.1	*Takifugu rubripes* transcription factor *Sox*2 (*Sox*2) gene, complete cds	93%
EF431922.1	*Oreochromis karongae Sox*2 (*Sox*2) gene, partial cds	99%
EF431923.1	*Oreochromis mossambicus Sox*2 (*Sox*2) gene, partial cds	99%
DQ900581.1	*Gadus morhua Sox*2 mRNA, partial cds	89%
AB242329.1	*Danio rerio Sox*2 mRNA for transcription factor *SoxSox*2, complete cds	82%
NM_001173447.1	*Felis catus* sex determining region Y-box 2 (*SoxSox*2) mRNA, complete cds	83%
XM_545216.2	PREDICTED: *Canis familiaris* similar to Transcription factor *SoxSox*-2 (LOC488092), mRNA	82%
XM_002807565.1	PREDICTED: *Callithrix jacchus* transcription factor *SoxSox*-2-like (LOC100407856), mRNA	82%
XM_002716451.1	PREDICTED: *Oryctolagus cuniculus* SRY (sex determining region Y)-box	82%
EU503117.1	*Sus scrofa* sex determining region Y-box 2 (*Sox*2) mRNA, complete cds	82%
NG_009080.1	*Homo sapiens* SRY (sex determining region Y)-box 2 (*SoxSox*2), RefSeqGene on chromosome 3	81%
XM_002814321.1	PREDICTED: *Pongo abelii* transcription factor *SoxSox*-2-like (LOC100460719), mRNA	82%
NM_001105463.1	*Bos taurus* SRY (sex determining region Y)-box 2 (*SoxSox*2), mRNA	81%
XM_001506934.1	PREDICTED: *Ornithorhynchus anatinus* similar to *Sox*2 transcription factor (LOC100075503), mRNA	81%
NM_001109181.1	*Rattus norvegicus* SRY (sex determining region Y)-box 2 (*Sox*2), mRNA	81%
NM_001142940.1	*Macaca mulatta* SRY (sex determining region Y)-box 2 (*SoxSox*2), mRNA	81%
XM_516895.2	PREDICTED: *Pan troglodytes* similar to *Sox*-2 (LOC460869), mRNA	83%
NM_011443.3	*Mus musculus* SRY-box containing gene 2 (*Sox*2), mRNA	81%
NM_001143799.1	*Equus caballus* SRY (sex determining region Y)-box 2 (*SoxSox*2), mRNA	80%
D50603.1	*Gallus gallus* mRNA for *SoxSox*-2, complete cds	79%
XM_001368783.1	PREDICTED: *Monodelphis domestica* similar to *Sox*2 transcription factor (LOC100014523), mRNA	85%
BC159121.1	*Xenopus tropicalis* SRY (sex determining region Y)-box 1, mRNA (cDNA clone MGC:172383 IMAGE:7663004), complete cds	86%

Supplementary material 2: Similarity of *Sox14* gene between *O. niloticus* and several vertebrates based on Blast analysis at NCBI.

Accession number	**Sequence**	**Match score**
EF431925.1	*Oreochromis aureus Sox*14 (*Sox*14) gene, complete cds	99%
EF431927.1	*Oreochromis mossambicus Sox*14 (*Sox*14) gene, partial CDs	99%
EF431926.1	*Oreochromis karongae Sox*14 (*Sox*14) gene, partial cds	99%
AY277955.1	*Takifugu rubripes* transcription factor *Sox*14a (*Sox*14a) gene, complete cds	90%
NM_001164872.1	*Oryzias latipes* SRY-box containing gene 14 (*Sox*14)	91%
NM_001194657.1	*Macaca mulatta* SRY (sex determining region Y)-box 14 (*SoxSox*14), mRNA	80%
AB063098.1	*Macaca fascicularis* brain cDNA clone:QtrA-14282, similar to human SRY-box 14 (*SoxSox*14), mRNA	80%
XM_526317.2	PREDICTED: Pan troglodytes SRY-box 14 (*SoxSox*14), mRNA	79%
NM_004189.2	*Homo sapiens* SRY (sex determining region Y)-box 14 (*SoxSox*14)	79%
NM_011440.1	*Mus musculus* SRY-box containing gene 14 (*Sox*14), mRNA	78%
NM_001163781.1	*Bos taurus* SRY (sex determining region Y)-box 14 (*SoxSox*14), mRNA	78%
XM_002742479.1	PREDICTED: *Callithrix jacchus* transcription factor *SoxSox*-21-like, transcript variant 1 (LOC100391037), mRNA	83%
XM_002814084.1	PREDICTED: *Pongo abelii* transcription factor *SoxSox*-14-like (LOC100460230), mRNA	79%
XM_001916428.1	PREDICTED: *Equus caballus* similar to rCG37201 (LOC100147540), mRNA	80%
AY612092.1	*Tetraodon nigroviridis* transcription factor *Sox* 8 (*SoxSox*8) mRNA, complete cds	100%

5. CONCLUSÕES

O grupo dos peixes é um excelente modelo para estudos evolutivos por conta da plasticidade da sua estrutura genômica evidenciada pela grande variação no tamanho dos genomas, números cromossômicos e presença de polimorfismos. Com a análise comparativa do genoma dos peixes com o genoma de outros vertebrados é possível inferir sobre a evolução de genes e grandes blocos genômicos em escala taxonômica dos grandes grupos de vertebrados.

Neste trabalho, utilizando dados disponíveis de sequenciamento do genoma e de bibliotecas em BACs da espécie *Oreochromis niloticus,* foi possível analisar com mais detalhes, pela integração de técnicas citogenéticas e de genômica comparativa, a evolução cromossômica da família Cichlidae, em especial no grupo africano Pseudocrenilabrinae. Em geral, o grupo aparece com seu genoma bastante conservado em relação a número diploide e localização dos marcadores/genes presentes nos BACs utilizados, mostrando a manutenção de grandes blocos genômicos equivalendo a cromossomos inteiros em diferentes espécies. Algumas variações observadas devem-se a rearranjos pontuais ocorridos, não afetando a estrutura genômica geral do grupo, mas indicando que o genoma dos ciclídeos continua sofrendo mudanças.

Ainda, quando comparamos o padrão encontrado para os ciclídeos com outros peixes e até mesmo outros vertebrados, podemos perceber a conservação de blocos de genes em diversos grupos. A organização geral do genoma pode ter sido rearranjada por milhões de anos por meio de quebras cromossômicas, fusões, duplicações e deleções; mas ainda é possível identificar blocos menores mantendo a mesma ordem dos genes.

Ao contrário do esperado, tomando por base dados que indicam a grande plasticidade genômica presente no genoma dos peixes em geral, observa-se uma grande conservação dos genomas das espécies de peixes estudadas, que incluem inclusive

sintenias conservadas com outros grupos de vertebrados. Isto pode ser reflexo das características funcionais da organização dos genomas, favorecendo a manutenção e a estabilidade da estrutura ou funções cromossômicas.

O mapa citogenético gerado para a família Cichlidae, integrado aos dados genômicos, abre portas para uma nova perspectiva nos estudos cromossômicos de peixes em geral, nos quais a integração de dados genômicos à citogenética molecular, permite uma análise mais apurada do genoma e da sua dinâmica evolutiva.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Aerts J, Crooijmans R, Cornelissen S, Hemmatian K, Veenendaal T, Jaadar A, van der Poel J, Fillon V, Vignal A, Groenen M. (2003) Integration of chicken genomic resources to enable whole-genome sequencing. Cytogenet Genome Res 102: 297-303.

Agresti JJ, Seki S, Cnaani A, Poompuang S, Hallerman EM, Umiel N, Hulata G, Gall GAE, May B. (2000) Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. Aquaculture 185: 43-56.

Ahn S, Tanksley SD. (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci, USA 90:7980-7984.

Albertson RC, Streelman JT, Kocher TD. (2003a) Genetic basis of adaptive shape differences in the cichlid head. J Heredity 94(4): 291-301.

Albertson RC, Streelman JT, Kocher TD. (2003b) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc Natl Acad Sci, USA 100: 5252-5257.

Axelrod HR. (1996) The most complete colored lexicon of cichlids. 2nd ed (TF Publications, NJ).

Barbazuk WB, Korf I, Kadavi C, Heyen J, Tate S, Wun E, Bedell JA, McPherson JD, Johnson SL. (2000) Syntenic relationship of the zebrafish and human genomes. Genome Res. 10: 1351–1358.

Barlow GW. (2000) The Cichlid's Fishes: Nature's Grand Experiment in Evolution. Perseus Publishing, Cambridge, MA.

Baroiller JF, D'Cotta H. (2001) Environment and sex determination in farmed fish. Comp Biochem Physiol C, 130: 399-409.

Basic Local Alignment Search Tool (BLAST) at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Beck TW, Menninger J, Voigt G, et al. (2001) Comparative feline genomics: a BAC/PAC contig map of the major histocompatibility complex class II region. Genomics 71: 282-295.

Becker SED, Thomas R, Trifonov VA, Wayne RK, Graphodatsky AS, Breen M. (2011) Anchoring the dog to its relatives reveals new evolutionary breakpoints across 11 species of the Canidae and provides new clues for the role of B chromosomes. Chromosome Res. 19:685–708.

Beldade P, Saenko SV, Pul N, Long AD. (2009) A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome. PLoS Genet 5: e1000366.

Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, Brown LG, Rozen S, Warren WC, Wilson RK, Page DC. (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 466:612–616.

Bertollo LAC, Takahashi CS, Moreira-Filho O. (1978) Cytotaxonomic consideration on *Hoplias lacerdae* (Pisces, Erythrinidae). Braz J Genet 1:103–120.

Bettin Pires L, Giuliano-Caetano L, Dias A L. (2010) Cytogenetic characterization of *Geophagus brasiliensis* and two species of *Gymnogeophagus* (Cichlidae: Geophaginae) from Guaíba Lake, RS, Brazil. Folia biol. (Kraków) 58: 29-34.

BouillaBase-Comparative Genome Browsers (www.BouillaBase.org).

Bowles J, Schepers G, Koopman P. (2000) Phylogeny of the *Sox* family of developmental transcription factors based on sequence and structural indicators. Dev Biol 227:239– 255.

Britski H.A. (1972) Peixes de água doce do estado de São Paulo: sistemática. In: Poluição e Piscicultura, Faculdade de Saúde Pública da USP- Instituto de Pesca da C.P.R.N. da Secretaria de Agricultura. 79:108.

Burt DW, Bruley CK, Dunn I, et al. (1999) Dynamics of chromosome evolution, clues from comparative gene mapping in birds and mammals. Nature 402:411–413.

Carleton KL, Hofmann CM, Klisz C, Patel Z, Chircus LM, Simenauer LH, Soodoo N, Albertson RC, Ser JR. (2010) Genetic basis of differential opsin gene expression in cichlid fishes. J Evol Biol. 23(4):840-53.

Carrasco LAP, Penman DJ, Bromage N. (1999) Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. Aquaculture 173: 207–218.

Chang X, Kobayashi T, Senthilkumaran B, Kobayashi-Kajura H, Sudhakumari CC, Nagahama Y. (2005) Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). Gen Comp Endocrinol. 141:101–15.

Chapus C, Edwards SV. (2009) Genome evolution in Reptilia: in silico chicken mapping of 12,000 BAC-end sequences from two reptiles and a basal bird. BMC Genomics. 10.

Charlesworth D, Mank JE. (2010) The Birds and the Bees and the Flowers and the Trees: Lessons from Genetic Mapping of Sex Determination in Plants and Animals. Genetics 186: 9–31.

Chen R, Sodergren E, Weinstock GM, Gibbs RA. (2004) Dynamic building of a BAC clone tiling path for the rat genome sequencing project. Genome Res. 14: 679-684.

Cheng ZK, Presting GG, Buell CR, Wing RA, Jiang JM. (2001) High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by

genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. Genetics 157:1749-1757.

Chew JSK, Oliveira C, Wright JM, Dobson MJ. (2002) Molecular and cytogenetic analysis of the telomeric (TTAGGG)n repetitive sequences in the Nile tilapia, *Oreochromis niloticus* (Teleostei: Cichlidae). Chromosoma 111:45-52.

Children's Hospital Oakland Research Institute – (http://bacpac.chori.org/home.htm).

Cichlid genome browser (http://cichlid.umd.edu/cgi-bin/gb2/gbrowse/stickleback/).

Cmap Viewer (http://dev.cichlid.umd.edu/cmap/).

Cnaani A, Kocher TD. (2008) Sex-linked markers and microsatellite locos duplication in the cichlid species *Oreochromis tanganicae*. Biol. Lett. 4:700–703.

Cnaani A, Lee BY, Ozouf-Costaz C, Bonillo C, Baroiller JF, D'Cotta H, Kocher TD. (2007) Mapping of *Sox*2 and *Sox14* in Tilapia (*Oreochromis spp*.). Sex Dev 1: 207–210.

Cnaani A, Lee B-Y, Zilberman N, Ozouf-Costaz C, Hulata G, Ron M, D'Hont A, Baroiller J-F, D'Cotta H, Penman DJ, Tomasino E, Coutanceau J-P, Pepey E, Shirak A, Kocher TD. (2008) Genetics of Sex Determination in Tilapiine Species. Sex Dev 1: 43-54.

Collignon J, Sockanathan S, Hacker A, Cohen-Tannoudji M, Norris D, Rastan S et al. (1996) A comparison of the properties of *Sox-3* with *Sry* and two related genes, *Sox-1* and *Sox-2*. Development 122: 509–520.

Conceição IC, Long AD, Gruber JD, Beldade P. (2011) Genomic Sequence around Butterfly Wing Development Genes: Annotation and Comparative Analysis. PLoS One 6(8).

d'Alençon E, Sezutsu H, Legeai F, Permal E, Bernard-Samain S, Gimenez S, Gagneur C, Cousserans F, Shimomura M, Brun-Barale A, Flutre T, Couloux A, East P, Gordon K, Mita K, Quesneville H, Fournier P, Feyereisen R. (2010) Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. Proc Natl Acad Sci, USA. 107(17):7680-5.

Dalrymple BP, Kirkness EF, Nefedov M, McWilliam S, Ratnakumar A, Barris W, Zhao S, Shetty J, Maddox JF, O'Grady M, Nicholas F, Crawford AM, Smith T, de Jong PJ, McEwan J, Oddy VH, Cockett NE. (2007) The International Sheep Genomics Consortium: Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. Genome Biol. 8:R152.

Dalton BE, Cronin TW, Marshall NJ, Carleton KL. (2010) The fish eye view: are cichlids conspicuous? J Exp Biol 213: 2243–2255.

Danley PD, Kocher TD. (2011) Speciation in rapidly diverging systems: lessons from Lake Malawi. Mol Ecol. 10:1075-1086.

D'Cotta H, Fostier A, Guiguen Y, Govoroun M, Baroiller JF. (2001) Aromatase plays a key role during normal and temperature-induced sex differentiation of tilapia *Oreochromis niloticus*. Mol Reprod Dev. 59:265–76.

Deutsch JC. (1997) Colour diversification in Malawi cichlids: evidence for adaptation, reinforcement or sexual selection?" Biological Journal of the Linnean Society 62: 1–14.

Devlin RH, Nagahama Y. (2002) Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208: 191-364.

Di Meo GP, Goldammer T, Perucatti A, Genualdo V, Iannuzzi A, Incarnato D, Rebl A, Di Berardino D, Iannuzzi L. (2011) Extended Cytogenetic Maps of Sheep

Chromosome 1 and Their Cattle and River Buffalo Homoeologues: Comparison with the OAR1 RH Map and Human Chromosomes 2, 3, 21 and 1q Cytogenet Genome Res 133: 16–24.

Di Palma F, Kidd C, Borowsky R, Kocher TD. (2007) Construction of bacterial artificial chromosome libraries for the Lake Malawi cichlid (*Metriaclima zebra*), and the blind cavefish (*Astyanax mexicanus*). Zebrafish 4:41-7.

Doligez A, Adam-Blondon AF, Cipriani G, Di Gaspero G, Laucou V, Merdinoglu D, Meredith CP, Riaz S, Roux C, This P. (2006) An integrated SSR map of grapevine base don five mapping populations. Theor Appl Genet 113: 369-382.

Dominey WJ. (1984) Effects of sexual selection and life history on speciation: species flocks in African cichlids and Hawaiian *Drosophila*. In: Evolution of Fish Species Flocks, Echelle AA, Kornfield I (Eds)., University of Maine at Orono Press, Orono,Me, USA, 231–250.

Duke SE, Samollow PB, Mauceli E, Lindblad-Toh K, Breen M. (2007) Integrated cytogenetic BAC map of the genome of the gray, short-tailed opossum, *Monodelphis domestica*. Chromosome Res, 15: 361–370.

Ellegren H. (2010) Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol 25: 283-291.

Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA. (2005) A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. Proc Natl Acad Sci, USA 102: 18526-31.

Ezaz MT, Harvey SC, Boonphakdee C, Teale AJ, McAndrew BJ, Penman DJ. (2004) Isolation and Physical Mapping of Sex-Linked AFLP Markers in Nile Tilapia (*Oreochromis niloticus L.*). Mar Biotechnol, 6:435–445.

FAO (2009) World aquaculture production of fish, crustaceans, mollusks, etc., by principal species in 2009. http://www.fao.org/corp/statistics/en/). Acessado em 20 de Novembro de 2011.

Farias IP, Ortí G, Sampaio I, Schneider H, Meyer A. (1999) Mitochondrial DNA phylogeny of the family Cichlidae: monophyly and fast molecular evolution of the Neotropical assemblage. J Mol Evol 48: 703–711.

Feldberg E, Bertollo LAC. (1984) Discordance in chromosome number among somatic and gonadal tissue cells of *Gymnogeophagus balzanii* (Pisces, Cichlidae). Rev Brasil Genet 7: 639-645.

Feldberg E, Bertollo LAC. (1985) Karyotypes of 10 species of Neotropical Cichlids (Pisces, Perciformes). Caryologia 38: 257-268.

Feldberg E, Porto JIR, Alves-Brinn MN, Mendonça MNC, Bezaquem DC. (2004) B chromosomes in Amazonia cichlid species. Cytogenet Genome Res, 106: 195-198.

Feldberg E, Porto JIR, Bertollo LAC. (2003) Chromosomal changes and adaptation of cichlid fishes during evolution. In: Fish adaptations. Val AL, Kapoor BG (Eds). Science Publishers, Inc. New Dehli & New York, 285:308, 418pp.

Ferguson-Smith MA, Trifonov V. (2007) Mammalian karyotype evolution. Nat Rev Genet 8: 950-962.

Ferreira IA, Martins C. (2008) Physical chromosome mapping of repetitive DNA sequences in Nile tilapia *Oreochromis niloticus*: evidences for a differential distribution of repetitive elements in the sex chromosomes. Micron 39: 411–418.

Ferreira IA, Poletto AB, Mota-Velasco J, Penman DJ, Kocher TD, Martins C. (2010) Chromosome evolution in African cichlid fish: contributions from the physical mapping of repeated DNAs. Cytogenet and Genome Res. 129: 314-322.

Foresti F, Oliveira C, Galetti Jr PM, Almeida-Toledo LF. (1993) Synaptonemal complex analysis in spermatocytes of tilapia, *Oreochromis niloticus* (Pisces, Cichlidae). Genome 36: 1124-1128.

Forster AC, McInnes JL, Skingle DC, Symons RH. (1985) Non-radioactive hybridization probes prepared by the chemical labelling of DNA and RNAwith a novel reagent, photobiotin. Nucleic Acids Res. 13: 745-761.

Freeman JL, Adeniyi A, Banerjee R, Dallaire S, Maguire SF, Chi J, Ng BL, et al (2007) Definition of the zebrafish genome using flow cytometry and cytogenetic mapping. BMC Genomics 8:195.

Genbank (http://www.ncbi.nlm.nih.gov/nuccore).

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.

Genomicus (http://www.dyogen.ens.fr/genomicus/).

Giovannotti M, Caputo V, O'Brien PCM, Lovell FL, Trifonov V, Nisi Cerioni P, Olmo E, Ferguson-Smith MA, Rens W. (2009) Skinks (Reptilia: Scincidae) have highly conserved karyotypes as revealed by chromosome painting. Cytogenet Genome Res 127:224–231.

Goldammer T, Brunner RM, Rebl A, et al. (2009) Cytogenetic anchoring of radiation hybrid and virtual maps of sheep chromosome X and comparison of X chromosomes in sheep, cattle, and human. Chromosome Research 17: 497–506.

Gong S, Zheng C, Doughty ML, et al. (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425: 917-925.

Graphodatsky AS, Trifonov VA, Stanyon R. (2011) The genome diversity and karyotype evolution of mammals. Mol Cytogenet 4:22.

Graphodatsky AS. (1989) Conserved and variable elements of mammalian chromosomes. In Cytogenetics of animals**.** CRE H. Oxon (Ed). UK: CAB International Press, 95-124.

Griffin DK, Robertson LB, Tempest HG, Skinner BM. (2007) The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenet Genome Res. 117: 64–77.

Groenen MAM, Cheng HH, Bumstead N, et al. (2000) A consensus linkage map of the chicken genome. Genome Res. 10: 137–147.

Gross MC, Schneider CH, Valente GT, Martins C, Feldberg E. (2010) Variability of 18S rDNA locos among *Symphysodon* fishes: chromosomal rearrangements. J Fish Biol. 76: 1117-27.

Gross MC, Schneider CH, Valente GT, Porto JI, Martins C, Feldberg E. (2009) Comparative cytogenetic analysis of the genus *Symphysodon* (discus fishes, cichlidae): chromosomal characteristics of retrotransposons and minor ribosomal DNA. Cytogenet Genome Res. 127: 43-53.

Grützner F, Crollius HR, Lütjens G, Jaillon O, Weissenbach J, Ropers HH, Haaf T. (2002) Four-Hundred Million Years of Conserved Synteny of Human Xp and Xq Genes on Three *Tetraodon* Chromosomes. Genome Res. 12: 1316-1322.

Grützner F, Lütjens G, Rovira C, Barnes DW, Ropers HH, Haaf T. (1999) Classical and molecular cytogenetics of the pufferfish *Tetraodon nigroviridis*. Chromosome Res**. 7:** 655–662.

Guo B, Tong C, He S. (2009) *Sox* genes evolution in closely related young tetraploid cyprinid fishes and their diploid relative. Gene 439: 102–112.

Guth SIE, Wegner M. (2008) Having it both ways: *Sox* protein function between conservation and innovation. Cell Mol Life Sci. 65: 3000 – 3018.

Hargrave M, James K, Nield K, Toomes C, Georgas K, Sullivan T, et al. (2000) Fine mapping of the neurally expressed gene *Sox14* to human 3q23, relative to three congenital diseases. Hum Genet 106: 432–439.

Harper LC, Cande WZ. (2000) Mapping a new frontier; development of integrated cytogenetic maps in plants. Funct Integr Genomics 1: 89-98.

Harvey SC, Boonphakdee C, Campos-Ramos R, et al. (2003) Analysis of repetitive DNA sequences in the sex chromosomes of *Oreochromis niloticus* . Cytogenet Genome Res. 101: 314–319.

Harvey SC, Campos-Ramos R, Kennedy RR, et al. (2002) Karyotype evolution in Tilapia: mitotic and meiotic chromosome analysis of *Oreochromis karongae* and *O. niloticus* × *O. karongae* hybrids. Genetica 115:169–177.

Hasterok R, Marasek A, Donnison IS, et al. (2006) Alignment of the Genomes of *Brachypodium distachyon* and Temperate Cereals and Grasses Using Bacterial Artificial Chromosome Landing With Fluorescence in Situ Hybridization. Genetics 173: 349–362.

Hope RM, Cooper S, Wainwright B. (1990) Globin macromolecular sequence in marsupials and monotremes. In: Mammals from pouches and eggs: genetic breeding and the evolution of marsupials and monotremes. Graves JAM, Hope RM, Cooper DW (Eds). CSIRO Press, Melbourne, 147–171.

Hoskins RA, Nelson CR, Berman BP, et al. (2000) A BAC-Based Physical Map of the Major Autosomes of *Drosophila melanogaster.* Science 287.

Hossain A, Saunders GF. (2001) The Human Sex-determining Gene *SRY* Is a Direct Target of *WT1**. J Biol Chem 276(20):16817–16823.

Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. (2006) Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. Dev Biol 300: 90–107.

Huan P, Zhang XJ, Zhang XJ, et al. (2009) Chromosomal localization and development of SNP markers of a serine protease gene in Farrer's scallop (*Chlamys farreri*). Yi Chuan 31: 1241-1247.

Instituto Sanger Ensembl Database (http : / www.ensembl.org /).

Iovene M, Cavagnaro PF, Senalik D, Buell CR, Jiang J, Simon PW. (2011) Comparative FISH mapping of *Daucus* species (Apiaceae family). Chromosome Res 19: 493–506.

Jaillon O, Aury J-M, Brunet F, et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431: 947-957.

Jiang J, Gill BS, Wang G-L, Ronald PC, Ward DC. (1995) Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci, USA 92: 4487-4491.

Jiang J, Gill BS. (2006) Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome 49: 1057-1068.

Kamachi Y, Uchikawa M, Kondoh H. (2000) Pairing *SOX* off with partners in the regulation of embryonic development. Trends Genet 16: 182– 187.

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: 200-4.

Katagiri T, Kidd C, Tomasino E, et al. (2005) A BAC-based physical map of the Nile tilapia genome. BMC Genomics 89: 1-6.

Kawai A, Nishida-Umehara C, Ishijima J, Tsuda Y, Ota H, Matsuda Y. (2007) Different origins of bird and reptile sex chromosomes inferred from comparative mapping of chicken Z linked genes. Cytogenet Genome Res. 117: 92-102.

Keller B, Feuillet C. (2000) Colinearity and gene density in grass genomes. Trends Plant Sci. 5(6): 246-51.

Kemkemer C, Kohn M, Cooper DN, Froenicke L, Högel J, Hameister H, Kehrer-Sawatzki H. (2009) Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. BMC Evol Biol 9:84.

Kiefer JC. (2007) Back to Basics: *Sox* Genes. Dev Dyn. 236: 2356–2366.

Kirby PJ, Waters PD, Delbridge M, Svartman M, Stewart AN. (2002) Cloning and mapping of platypus *Sox*2 and *Sox*14: insights into *Sox* group B evolution. Cytogenet Genome Res. 98: 96–100.

Kishi M, Mizuseki K, Sasai N, Yamazaki H, Shiota K, Nakanishi S, et al. (2000) Requirement of *Sox*2-mediated signaling for differentiation of early *Xenopus* neuroectoderm. Development 127: 791–800.

Klein PE, Klein RR, Cartinhour SW, et al. (2000) A high-throughput AFLP base method for constructing integrated genetic and physical maps: Progress towards a sorghum genome map. Genome Res. 10: 789–807.

Kobayashi N, Watanabe M, Horiike T, Kohara Y, Okada N. (2009) Extensive analysis of EST sequences reveals that all cichlid species in Lake Victoria share almost identical transcript sets. Gene 441: 187-91.

Kocher TD, Lee WJ, Sobolewska H, Penman D, McAndrew B. (1998) A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). Genetics 148: 1225-1232.

Kocher TD. (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet. 5: 288-298.

Kohn M, Högel J, Vogel W, et al. (2006) Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends Genet. 22: 203–210.

Koopman P, Schepers G, Brenner S, Venkatesh B. (2004) Origin and diversity of the *Sox* transcription factor gene family: Genome-wide analysis in *Fugu rubripes*. Gene 328: 177–186.

Kornfield IL. (1984) Descriptive genetics of cichlids fishes. In: Evolutionary genetics of fishes. Turner BJ (Ed). Press, New York, 591-616.

Kosswig C. (1964) Polygenic sex determination. Experientia 20: 190-199.

Kuhl H, Sarropoulou E, Tine M, Kotoulas G, Magoulas A, Reinhardt R. (2011) A Comparative BACMap for the Gilthead Sea Bream (*Sparus aurata L.*). J Biomed Biotechnol :329025.

Kullander SO. (2003) Family Cichlidae. In: Check list of the freshwater fishes of South and Central America. Reis RE, Kullander SO, Ferraris CJ. EDIPUCRS, RS, 605-654.

Kwon JY, McAndrew BJ, Penman DJ. (2001) Cloning of brain aromatase gene and expression of brain and ovarian aromatase genes during sexual differentiation in genetic male and female Nile tilapia *Oreochromis niloticus***.** Mol Reprod Dev. 59: 359–70.

Lander ES, Linton LM, Birren B, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Lang M, Miyake T, Braasch I, Tinnemore D, Siegel N, Salzburger W, Amemiya CT, Meyer A. (2006) A BAC library of the East African haplochromine cichlid fish *Astatotilapia burtoni*. J Exp Zool B Mol Dev Evol. 306: 35-44.

Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM. (2008) Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol. 25: 980–996.

Lee BY, Coutanceau JP, Ozouf-Costaz C, D'Cotta H, Baroiller JF, Kocher T. (2011) Genetic and Physical Mapping of Sex Linked AFLP Markers in Nile Tilapia (*Oreochromis niloticus*). Mar Biotechnol. 13: 557–562.

Lee BY, Howe AE, Conte MA, D'Cotta H, Pepey E, Baroiller JF, di Palma F, Carleton KL, Kocher TD. (2010) An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. BMC Genomics 11: 278.

Lee B-Y, Hulata G, Kocher TD. (2004) Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). Heredity 92: 543-549.

Lee BY, Kocher TD. (2007a) Comparative genomics and positional cloning. In: Aquaculture Genome Technologies. Liu J. (ed.). Blackwell, 325-337.

Lee BY, Kocher TD. (2007b) Exclusion of *Wilms tumor* (*WT1b*) and ovarian *cytochrome P450 aromatase* (*CYP19A1*) as candidates for sex determination genes in Nile tilapia (*Oreochromis niloticus*). Anim Genet. 38: 85-86.

Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD. (2005) A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). Genetics 170: 237-44.

Lee B-Y, Penman DJ, Kocher TD. (2003) Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. Anim Genet 34(5): 379-383

Levtchenko EN, Monnens LA, Bökenkamp A, Knoers NV. (2007) From gene to disease; Dent's disease caused by abnormalities in the *CLCN5* and *OCRL1* genes. Ned Tijdschr Geneeskd. 151: 2377-80.

Li R, Mignot E, Faraco J, Kadotani H, Cantanese J, Zhao B, Lin X, Hinton L, Ostrander EA, Patterson DF, Jong PJ. (1999) Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library. Genomics 58: 9-17.

Liem KF. (1991) Functional morphology. In: Cichlid fishes: behavior, ecology and evolution. Keenleyside MHA (Ed). Chapmam and Hall, London: 129 – 150.

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.

Lombarrd V, Delourne R. (2001) A consensus linkage map for repressed (*Brassica napus* L): construction and integration of their individual maps from DH populations. Theor Appl Genet. 103: 491-507.

Lowe-McConnell RH. (1991) Ecology of cichlids in South American and African waters, excluding the African Great Lakes. In: Cichlid fishes: behaviour, ecology and evolution. Keenleyside MHA. (Ed.). Chapman and Hall, London, 60-85.

Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent WJ, Blanchette M, Haussler D, Miller W. (2006) Reconstructing contiguous regions of an ancestral genome. Genome Res. 16: 1557-1565.

Majumdar KC, McAndrew BJ. (1986) Relative DNA content of somatic nuclei and chromosomal studies in three genera, *Tilapia*, *Sarotherodon*, and *Oreochromis* of the tribe Tilapiini (Pisces, Cichlidae). Genetica 68: 175–188.

Mank JE, Promislow DEL, Avise JC. (2006) Evolution of alternative sex-determining mechanisms in teleost fishes. Biol. J. Linnean Soc. 87: 83-93.

Martins C, Cabral-de-Mello DC, Valente GT, Mazzuchelli J, Oliveira SG, Pinhal D. (2011) Animal genomes under the focus of cytogenetics. New Hampshire: Nova Science Publisher.

Martins C, Oliveira C, Wasko AP, Wright JM. (2004) Physical mapping of the Nile tilapia (*Oreochromis niloticus*) genome by fluorescent in situ hybridization of repetitive DNAs to metaphase chromosomes - A review. Aquaculture 231: 37-49.

Martins C, Wasko AP, Oliveira C, Porto-Foresti F, Parise-Maltempi PP, Wright JM, Foresti F. (2002) Dynamics of 5S rDNA in the tilapia (*Oreochromis niloticus*) genome: Repeat units, inverted sequences, pseudogenes and chromosome loci. Cytogenet Genome Res 98: 76.85.

Martins C, Wasko AP, Oliveira C, Wright J. (2000) Nucleotide sequence of 5S rDNA and localization of the ribosomal RNA genes to metaphase chromosomes of the Tilapiini cichlid fish, *Oreochromis niloticus*. Hereditas 133(1): 39-46.

Martins-Santos IC, Portela-Castro ALB, Júlio jr HF. (1995) Chromosomes analysis of 5 species of the Cichlidae family (Pisces, Perciformes) from the Parana river. Cytologia 60: 223-231.

Mazzuchelli J, Martins C. (2009) Genomic Organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. Genetica 136: 461-9.

Mazzuchelli J, Yang F, Kocher TD, Martins C. (2011) Comparative cytogenetic mapping of *SOX2* and *SOX14* in cichlid fishes and inferences on the genomic organization of both genes in vertebrates. Chromosome Res 19: 657–667.

McElroy DM, Kornfield I, Everett J. (1991) Coloration in African cichlids: diversity and constraints in Lake Malawi endemics," Netherlands J of Zoology 41: 250–268.

McKimmie C, Woerfel G, Russell S. (2005) Conserved genomic organization of Group B *Sox* genes in insects. BMC Genet 19: 6-26.

Meulemans D, Bronner-Fraser M. (2007) The amphioxus *SOX*B family: implications for the evolution of vertebrate placodes. Int J Biol Sci 3: 356–364.

Mitani H, Kamei Y, Fukamachi S, Oda S, Sasaki T, Asakawa S, Todo T, Shimizu N. (2006) The medaka genome: why we need multiple fish models in vertebrate functional genomics. Genome Dyn 2: 165-82.

Miyake T, Amemiya CT. (2004) BAC libraries and comparative genomics of aquatic chordate species. Comp Biochem Physiol. 138: 233-244.

Mota-Velasco JC, Ferreira IA, Cioffi MB, Ocalewicz K, Campos-Ramos R, Shirak A, Lee BY, Martins C, Penman DJ. (2010) Characterization of the chromosome fusions in *Oreochromis karongae.* Chromosome Res 18: 575–586.

Murphy WJ, Larkin DM, Wind AE van der, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'brien SJ, Pevzner PA, Lewin HA. (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309: 613-617.

Murray AM. (2001) The fossil record and biogeography of the Cichlidae (Actinopterygii, Labroidei). Biol J Linnean Society 74: 517-532.

Nakatani Y, Takeda H, Kohara Y, Morishita S. (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 17: 1254–1265.

Nanda I, Benisch P, Fetting D, Haaf T, Schmid M. (2011) Synteny conservation of chicken macrochromosomes 1-10 in different avian lineages revealed by cross-species chromosome painting. Cytogenet Genome Res. 132(3):165-81.

Nanda I, Haaf T, Schartl M, Schmid M, Burt DW. (2002) Comparative mapping of Z-orthologous genes in vertebrates: implications for the evolution of avian sex chromosomes. Cytogenet Genome Res. 99:178–184.

Nanda I, Shan Z, Schartl M, Burt DW, Koehler M, Nothwang H-G, Grützner F, Paton IR, Windsor D, Dunn I, Engel W, Staeheli P, Mizuno S, Haaf T, Schmid M. (1999) 300 million years of conserved synteny between chicken Z and human chromosome 9. Nat Genet 21: 258-259.

Nanda I, Zend-Ajusch E, Shan Z, Grützner F, Schartl M, Burt DW, Koehler M, Fowler VM, Goodwin G, Schneider WJ, Mizuno S, Dechant T, Haaf T, Schmid M. (2000) Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene DMRT1: a comparative (re)view on avian sex determination. Cytogenet Cell Genet 89: 67-78.

National Center for Biotechnology Information (NCBI) databases (www.ncbi.nlm.nih.gov/genbank/).

National Human Genome research Institute – (http://www.genome.gov).

NCBI Map Viewer (http: //www.ncbi.nlm.nih.gov/genomes).

Nelson JS (2006) Fishes of the world. 4a. edição. J. Wiley & Sons., Inc. Hoboken. 601 pp.

O'Brien SJ, Menninger JC. (2006) Atlas of Mammalian Chromosomes. John Wiley and Sons Publishers.

Ocalewicz K, Mota-Velasco JC, Campos-Ramos R, Penman DJ. (2009) FISH and DAPI staining of the synaptonemal complex of the Nile tilapia (*Oreochromis niloticus*) allow orientation of the unpaired region of bivalent 1 observed during early pachytene. Chromosome Res. 17: 773–782.

Okuda Y, Yoda H, Uchikawa M, Furutani-Seiki M, Takeda H, Kondoh H, Kamachi Y (2006) Comparative genomic and expression analysis of group B1 *Sox* genes in zebrafish indicates their diversification during vertebrate evolution. Dev Dyn. 235: 811–825.

Oliveira C, Chew JSK, Porto-Foresti F, Dobson MJ, Wright JM. (1999) A LINE-like repetitive DNA sequence from the cichlid fish, *Oreochromis niloticus*: sequence analysis and chromosomal distribution. Chromosoma 108: 457-468.

Oliveira C, Foresti F, Hilsdorf AWS. (2009) Genetics of neotropical fish: from chromosomes to populations. Fish Physiol Biochem. 35: 81-100.

Oliveira C, Wang Y, Bryden LJ, Wright JM. (2003) Short interspersed repetitive elements (SINEs) from the cichlid fish, *Oreochromis niloticus*, and their chromosomal localization by fluorescent *in situ* hybridization. Caryologia 56: 181-189.

Oliveira C, Wright JM. (1998) Molecular cytogenetic analysis of heterochromatin in the chromosomes of tilapia, *Oreochromis niloticus* (Teleostei: Cichlidae). Chromosome Res. 6: 205-211.

Palti Y, Genet C, Gao G, Hu Y, You FM, Boussaha M, Rexroad CE 3rd, Luo MC. (2011) A Second Generation Integrated Map of the Rainbow Trout (*Oncorhynchus mykiss*) Genome: Analysis of Conserved Synteny with Model Fish Genomes. Mar Biotechnol (NY).

Paran I, Van der Voort JR, Lefebvre V, et al. (2004) A integrated genetic linkage map of pipper (*Capsicum* spp). Mol Breed 13: 251-261.

Pauers MJ, Kapfer JM, Fendos CE, Berg CS. (2008) Aggressive biases towards similarly coloured males in Lake Malawi cichlid fishes. Biol Lett, 4(2): 156– 159.

Pauers MJ. (2011) One Fish, Two Fish, Red Fish, Blue Fish: Geography, Ecology, Sympatry, and Male Coloration in the Lake Malawi Cichlid Genus *Labeotropheus* (Perciformes: Cichlidae). Int J Evol Biol, 575469.

Pedrosa A, Vallejos CE, Bachmair A, Schweizer D. (2003) Integration of common bean (*Phaseolus vulgaris L.*) linkage and chromosomal maps. Theor Appl Genet 106: 205-12.

Phillips RB, Keatley KA, Morasch MR, Ventura AB, Lubieniecki KP, Koop BF, Danzmann RG, Davidson WS. (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 Genetics 10: 46.

Pinkel D, Straume T, Gray JW. (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci, USA, 83: 2934-2938.

Pitnam WC, Cande S, LaBrecque J, Pindell J. (1993) Fragmentation of Gondwana: the separation of Africa from South America. In: Biology relationship between Africa and South America. Goldblatt P (Ed). Yale University Press, New Haven, 15-34.

Pokorná M, Giovannotti M, Kratochvíl L, Kasai F, Trifonov VA, O'Brien PCM, Caputo V, et al. (2011) Strong conservation of the bird Z chromosome in reptilian genomes is revealed by comparative painting despite 275 million years divergence. Chromosoma 120:455–468.

Poletto AB, Ferreira IA, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, Ribeiro HB, Venere PC, Nirchio M, Kocher TD, Martins C. (2010a) Chromosome differentiation patterns during cichlid fish evolution. BMC Genet. 11:50.

Poletto AB, Ferreira IA, Martins C. (2010b) The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 18S rRNA gene copies. BMC Genet. 11:1.

Popovic J, Stevanovic M. (2009) Remarkable evolutionary conservation of *SOX*14 orthologues. J Genet. 88: 15-24.

Price DJ. (1984) Genetic of sex determination in fishes – a brief review. In: Fish Reproduction: Strategies and Tactics. Potts GH, Wootton RJ (Eds.). London: Academic Press, 77-89.

Pringle EG, Baxter SW, Webster CL, Papanicolaou A, Lee SF, et al. (2007) Synteny and chromosome evolution in the Lepidoptera: Evidence from mapping in *Heliconius melpomene*. Genetics 177: 417–426.

Putnam NH, Srivastava M, Hellsten U, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317: 86-94.

Quiniou SMA, Katagiri T, Miller NW, Wilson M, Wolters WR, Waldbieser GC. (2003) Construction and characterization of a BAC library from a gynogenetic channel catfish, *Ictalurus punctatus*. Genet Sel Evol 35:1-11.

Ravi V, Venkatesh B. (2008) Rapidly evolving fish genomes and teleost diversity. Curr Opin Genet Dev 18: 544-50.

Reed KM, Chaves LD, Mendoza KM. (2007) An integrated and comparative genetic map of the turkey genome. Cytogenet Genome Res 119: 113-26.

Renn SC, Aubin-Horth N, Hofmann HA. (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. BMC Genomics 5: 42.

Revanna KV, Chiu C-C, Bierschank E, Dong Q. (2011) GSV: a web-based genome synteny viewer for customized dat. BMC Bioinformatics 12: 316.

Rexroad CE, Palti Y, Gahr SA, Vallejo RL. (2008) A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). BMC Genet 9: 74.

Robinson-Rechavi M, Marchand O, Escriva H, Bardet P, Zelus D, Hughes S, Laudet V. (2001a) Euteleost fish genomes are characterized by expansion of gene families. Genome Res 11: 781-788.

Robinson-Rechavi M, Marchand O, Escriva H, Laudet V. (2001b) An ancient whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. Curr Biol. 11: 458-459.

Romanov MN, Daniels LM, Dodgson JB, Delany ME. (2005) Integration of the cytogenetic and physical maps of chicken chromosome 17. Chromosome Res. 13: 215-222.

Sahara K, Yoshido A, Marec F, Fukova I, Zhang HB, Wu CC, Goldsmith MR, Yasukochi Y. (2007) Conserved synteny of genes between chromosome 15 of *Bombyx mori* and a chromosome of *Manduca sexta* shown by five-color BAC-FISH. Genome 50: 1061-1065.

Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. (2002) Phylogeny of the lake Tanganyika cichlid species flock and its relationship to the central and east African haplochromine cichlid fish faunas. Syst Biol. 51:113-135.

Salzburger W, Meyer A. (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. Naturwissenschaften 91: 277-290.

Salzburger W, Renn SC, Steinke D, Braasch I, Hofmann HA, Meyer A. (2008) Annotation of expressed sequence tags for the East African cichlid fish *Astatotilapia burtoni* and evolutionary analyses of cichlid ORFs. BMC Genomics 9:96.

Sanetra M, Henning F, Fukamachi S, Meyer A. (2009) A microsatellite-based genetic linkage map of the cichlid fish, *Astatotilapia burtoni* (Teleostei): a comparison of genomic architectures among rapidly speciating cichlids. Genetics 182: 387-97.

Sanger Institute Ensembl Database (http://www.ensembl.org).

Sarropoulou E, Nousdili D, Magoulas A, Kotoulas G. (2008) Linking the genomes of nonmodel teleosts through comparative genomics. Mar Biotechnol. 10(3): 227-33.

Satou Y, Satoh N. (2005) Cataloging transcription factor and major signaling molecule genes for functional genomic studies in *Ciona intestinalis*. Dev Genes Evol. 215: 580–596.

Schmid M, Nanda I, Guttenbach M, Steinlein C, Hoehn H, et al. (2000) First report on chicken genes and chromosomes 2000. Cytogenet Cell Genet. 90: 169–218.

Schwarzacher T. (2003) DNA, chromosomes, and in situ hybridization. Genome 46: 953–962.

Seehausen O, Alphen JJM van. (1999) Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlids diversity in Lake Victoria? Ecol Lett. 2: 262-271.

Seehausen O, Mayhew PJ, van Alphen JJM. (1999) Evolution of colour patterns in East African cichlid fish. J Evol Biol 12 (3): 514–534.

Seehausen O, Terai Y, Magalhaes IS et al. (2008) Speciation through sensory drive in cichlid fish. Nature 455: 620–626.

Seehausen O, Witte F, van Alphen JJM, Bouton N. (1998) Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. J Fish Biol 53: 37–55.

Ser JR, Roberts RB, Kocher TD. (2009) Multiple interacting loci control sex determination in Lake Malawi Cichlid fish. Evolution 64: 486-501.

Smith WL, Chakrabarty P, Sparks JS. (2008) Phylogeny, taxonomy, and evolution of Neotropical cichlids (Teleostei: Cichlidae: Cichlinae). Cladistics 24: 625-641.

Soler L, Conte MA, Katagiri T, Howe AE, Lee BY, Amemiya C, Stuart A, Dossat C, Poulain J, Johnson J, Di Palma F, Lindblad-Toh K, Baroiller JF, D'Cotta H, Ozouf-

Costaz C, Kocher TD. (2010) Comparative physical maps derived from BAC end sequences of tilapia (*Oreochromis niloticus*). BMC Genomics 11:636.

Sparks JS, Smith WL. (2004) Phylogeny and biogeography of cichlid fishes (Teleostei, Perciformes, Cichlidae). Cladistics 20: 501–517.

Srikulnath K, Nishida C, Matsubara K, Uno Y, Thongpan A, Suputtitada S, Apisitwanich S, Matsuda Y. (2009) Karyotypic evolution in squamate reptiles: comparative gene mapping revealed highly conserved linkage homology between the butterfly lizard (*Leiolepis reevesii rubritaeniata*, Agamidae, Lacertilia) and the Japanese four-striped rat snake (*Elaphe quadrivirgata*, Colubridae, Serpentes). Chromosome Res. 17: 975–986.

Sterba G. (1973) Freshwater fishes of the world. T.F.H. Publications, USA v1-2, 877p.

Stiglec R, Ezaz T, Graves JAM. (2007) Reassignment of chicken W chromosome sequences to the Z chromosome by fluorescence *in situ* hybridization (FISH). Cytogenet Genome Res. 116: 132–134.

Storey M, Mahoney JJ, Saunders AD, Duncan RA, Kelley SP, Coffin MF. (1995) Timing of hot-spot related volcanism and breakup of Madagascar and India. Science 167: 852-855.

Sumner AT. (1972) A simple technique for demonstrating centromeric heterochromatin. Exp. Cell. Res. 75: 304-306.

Sumner AT. (1990) Chromosome banding. Unwin Hyman, London: Unwin Hyman.

Tang X, Boer JM, Eck HJ, Bachem C, Visser RGF, Jong H. (2009) Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. Chromosome Res. 17: 899–915.

Teixeira WG, Ferreira IA, Cabral-de-Mello DC, Mazzuchelli J, Valente GT, Poletto AB, Venere PC, Martins C. (2009) Organization of repeated DNA elements in the

genome of the cichlid fish *Cichla kelberi* and its contributions to the knowledge of fish genomes. Cytogenet Genome Res. 125: 224-234.

The International Cichlid Genome Consortium. (2006) Genetic basis of vertebrate diversity: the cichlid fish model.

Thompson KW. (1979) Cytotaxonomy of 41 species of Neotropical Cichlidae. Copeia 679-691.

Thompson KW. (1981) Karyotypes of six species of African Cichlidae (Pisces: Perciformes). Experientia 37: 351–352.

Thorsen J, Zhu B, Frengen E, Osegawa K, de Jong PJ, et al. (2005) A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics 6:50.

Tunster SJ, Van De Pette M, John RM. (2011) BACs as Tools for the Study of Genomic Imprinting. J Biomed Biotechnol. 283013.

Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL. (2001) How many species of cichlid fishes are there in African lakes? Mol Ecol 10: 793-806.

Turner GF. (2007) Adaptive radiation of cichlid fish. Current Biology 17: 827-31.

Uchikawa M, Kamachi Y, Kondoh H. (1999) Two distinct group B *SoxSox* genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. Mech Dev. 84: 103–120.

UCSC Genome Bioinformatics – (http://genome.ucsc.edu/).

Ulloa M, Meridith WR, Shappley ZW, Kaphler AL. (2002) RFLP genetic linkage maps from F2.3 populations and joinmap of *Gossypium hirsutum*. Theor Appl Genet. 104: 200-208.

Valente GT, Mazzuchelli J, Ferreira IA, Poletto AB, Fantinatti BEA, Martins C. (2011) Cytogenetic Mapping of the Retroelements Rex1, Rex3 and Rex6 among Cichlid Fish: New Insights on the Chromosomal Distribution of Transposable Elements. Cytogenet and Genome Res. 133: 34-42.

Valente GT, Schneider CH, Gross MC, Feldberg E, Martins C (2009) Comparative cytogenetics of cichlid fishes through genomic in-situ hybridization (GISH) with emphasis on *Oreochromis niloticus*. Chromosome Res. 17: 791-799.

Venkatesh B. (2003) Evolution and diversity of fish genomes. Curr Opin Genet Dev 13: 588–592.

Vicari MR, Artoni RF, Moreira-Filho O, Bertollo LAC. (2006) Basic and molecular cytogenetics in freshwater Cichlidae (Osteichthyes, Perciformes). Karyotypic conservationism and divergence. Caryologia 59(3): 260-266.

Wang S, Xu P, Thorsen J, Zhu B, de Jong PJ, Waldbieser G, Kucuktas H, Liu Z. (2007) Characterization of a BAC library from channel catfish *Ictalurus punctatus*: indications of high levels of chromosomal reshuffling among teleost genomes. Mar Biotechnol. 9: 701-711.

Watanabe A, Shiina T, Shimizu S, Hosomichi K, Yanagiya K, Kita YF, Kimura T, Soeda E, Torii R, Ogasawara K, Kulski JK, Inoko H. (2007) A BAC-based contig map of the cynomolgus macaque (*Macaca fascicularis*) major histocompatibility complex genomic region. Genomics 89: 402–412.

Watanabe M, Kobayashi N, Fujiyama A, Okada N. (2003) Construction of a BAC library for *Haplochromis* chilotes, a cichlid fish from Lake Victoria. Genes Genet Syst. 78: 103-5.

Watanabe M, Kobayashi N, Shin-i T, Horiike T, Tateno Y, Kohara Y, Okada N. (2004) Extensive analysis of ORF sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock: identity of EST sequences between *Haplochromis chilotes* and *Haplochromis* sp. "Redtailsheller". Gene 343: 263-9.

Wei L, Cheng D, Li D, Meng M, Peng L, Tang L, Pan M, Xiang Z, Xia Q, Lu C. (2011) Identification and characterization of *Sox* genes in the silkworm, *Bombyx mori*. Mol Biol Rep. 38: 3573–3584.

Werner P, Mellersh CS, Raducha MG, DeRose S, Acland GM, Prociuk U, Wiegand N, Aguirre GD, Henthorn PS, Patterson DF, et al. (1999) Anchoring of canine linkage groups with chromosome-specific markers. Mamm. Genome 10: 814-823.

Wienberg J. (2004) The evolution of eutherian chromosomes. Curr Op Genet Devel. 14: 657-666.

Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan YL, Huang H, Postlethwait JH, Talbot WS. (2000) A comparative map of the zebrafish genome. Genome Res 10: 1903–1914.

Wu C, Proestou D, Carter D, Nicholson E, Santos F, Zhao S, Zhang H-B, Goldsmith M R. (2009) Construction and sequence sampling of deep-coverage, large-insert BAC libraries for three model lepidopteran species. BMC Genomics 10: 283.

Wu CC, Nimmakayala P, Santos FA, Springman R, Scheuring C, Meksem K, Lightfoot DA, Zhang HB. (2004) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet. 109: 1041–1050.

Yan Z, Denneboom C, Hatterndorf A, Dolstra O, Debener T, Stam P, Visser PB. (2005) Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers. Theoret Appl Genet 110: 766-777.

Yang F, Alkalaeva EZ, Perelman PL, Pardini AT, Harrison WR, O'Brien PC, Fu B, Graphodatsky AS, Ferguson-Smith MA, Robinson TJ. (2003) Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype. Proc Natl Acad Sci, USA 100: 1062-1066.

Yang F, Graphodatsky AS, Li T, Fu B, Dobigny G, Wang J, Perelman PL, Serdukova NA, Su W, O'Brien PC, Wang Y, Ferguson-Smith MA, Volobouev V, Nie W. (2006) Comparative genome maps of the pangolin, hedgehog, sloth, anteater and human revealed by cross-species chromosome painting: further insight into the ancestral karyotype and genome evolution of eutherian mammals. Chromosome Res. 14: 283-296.

Yasukochi Y, Ashakumary L, Baba K, Yoshido A, Sahara K. (2006) A second generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. Genetics 173: 1319–1328.

Yasukochi Y, Tanaka-Okuyama M, Shibata F, Yoshido A, Marec F, Wu C, Zhang H, Goldsmith MR, Sahara K. (2009) Extensive conserved synteny of genes between the karyotypes of *Manduca sexta* and *Bombyx mori* revealed by BAC-FISH mapping. PLoS One 4: e7465.

Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe M, Kuroiwa A, et al. (2011) B Chromosomes Have a Functional Effect on Female Sex Determination in Lake Victoria Cichlid Fishes. PLoS Genetics 7:e1002203.