RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo deste trabalho será disponibilizado somente a partir de 11/08/2019.

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

Genome assembly of the cichlid fish A*statotilapia latifasciata* with focus in population genomics of B chromosome polymorphism

Maryam Jehangir

Botucatu, July, 2017

UNIVERSIDADE ESTADUAL PAULISTA

“Julio de Mesquita Filho”

INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

GENOME ASSEMBLY OF THE CICHLID FISH *ASTATOTILAPIA LATIFASCIATA* WITH FOCUS IN POPULATION GENOMICS OF B CHROMOSOME POLYMORPHISM

MARYAM JEHANGIR

DR. CESAR MARTINS

Dissertation presented to the Institute of Biosciences, Campus of Botucatu, UNESP, to obtain the title of Master in the Postgraduate Program -University graduate in Biological Sciences (Genetics).

Botucatu- SP

2017

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Jehangir, Maryam.
Genome assembly of the cichlid fish astatotilapia latifasciata with focus in population genomics of B chromosome polymorphism / Maryam Jehangir. - Botucatu, 2017

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Cesar Martins
Coorientador: Guilherme Targino Valente
Capes: 20204000

1. Peixe - Genética. 2. Ciclídeos. 3. Cromossomos - Polimorfismo. 4. Genoma. 5. Genômica. 6. Citogenética.

Palavras-chave: Cromossomos B; Evolução; Montagem do genoma; Peixe cíclido; Polimorfismo.

Acknowledgements

In the name of Allah, the Most Gracious and the Most Merciful, for the strengths and His blessings in completing this thesis.

I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I thank my academic advisor, Professor Cesar Martins, for accepting me into his group. During my tenure, he gave me intellectual freedom in my work, supporting my attendance at various conferences, engaging me in new ideas, and demanding a high quality of work in all my endeavors. My appreciation to my co-advisor professor Guilherme Targino Valente, for beneficial ideas and knowledge regarding my research project.

Additionally, I would like to thank my committee members for their interest in my work.

Every result described in this thesis was accomplished with the help and support of fellow lab mates and collaborators especially, Syed Farhan Ahmad, Erica Ramos and Adauto Lima Cardoso. I greatly benefited from their keen scientific insight, their knack for solving seemingly intractable practical difficulties, and their ability to put complex ideas into simple terms.

I am indebted to UNESP administration, for providing much needed assistance with administrative tasks, reminding us of impending deadlines, and keeping our work running smoothly.

I am grateful for the FAPESP (process number: 2014/17683-6), which provides me research resources to pursue my master project.

I would like to acknowledge the Department of Genetics for arrangement of different courses and and the high-quality seminars, benefited greatly from these.

I must express my gratitude to Syed Farhan Ahmad, my husband, for his continued support, encouragement and collaborating in my research. Finally, I would like to acknowledge my mother, father, sisters and brother for their love and prayers. Also not forgetting my daughter, Inshirah Farhan, her cute smile gives me more courage.

Maryam Jehangir

(Master student)

Contents list

Abstract

B chromosomes (Bs) are additional to the standard regular chromosome set (As), and present in all groups of eukaryotes. A reference genome is key to understand genomics aspects of an organism. Here, we present the *de novo* genome assembly of the cichlid fish *A. latifasciata*: a well known model to study Bs. The assembly of *A. latifasciata* genome has not been performed so far. The main focus of this study is to analyze and assemble the *A. latifasciata* genome with no B (B-) and with B (B+) chromosomes. The assembled draft B- and B+ genomes comprised of 774 Mb and 781 Mb with 1.8 Mb and 2.5Mb of N50 value of scaffolds respectively, and spanning 23,391 number of genes. High coverage data with Illumina sequencing was obtained for males and females with 0B, 1B and 2B chromosomes to provide information regarding the population polymorphism of these genomes. We observed a high scale genomic diversity in all analyzed genomes showing a high rate/frequency of population polymorphism with no evident effect of B chromosome presence. However, the B specific single nucleotide polymorphisms were found in the sequences that were located on B chromosome. While, the whole-genome rearrangements (inter chromosomal translocations) were detected in B+ genome, and structural variations including insertions, deletions, inversions and duplications were predicted in a representative genomic region of B chromosome. These results bring an evidence that existence of Bs in a genome should favour the accumulations of mutations and structural polymorphisms in the amlified genomic regions present on B chromosmes. In addition, we also performed the coverage based sequence study coupled with FISH mapping which revealed: 1) the existence of high copy number of inactive Indian Hedgehog b (*Ihhb*) gene on B chromosome emerging as a pseudogene after series of duplication events ultimately becoming a major structural component of B; 2) B chromosome have incorporated the entire 45S RNA cluster (18S ribosomal RNA, internal transcribed spacer 1, 5.8S ribosomal RNA, internal transcribed spacer 2, and 28S ribosomal RNA) from the A complement set. The assembly of *A. latifasciata* genome will serve as a reference for genetic analysis and the approach presented in this paper opens the perspective to advance understanding B chromosomes biology.

Keywords: Genome Assembly, Cichlid fish, B chromosome, Genome, Sequencing, polymorphism, evolution

1. Introduction

1.1. B chromosomes

B chromosomes (Bs) were first reported more than a century ago by E. B. Wilson in the leaf-footed plant bug insect (*Metapodius*) (Wilson, 1907). Bs are accessories to the standard regular chromosome set (As) and also named as extra chromosomes that are present in some individuals of species. B chromosomes behave different from normal set of chromosomes because they do not pair or undergo recombination with the A chromosomes during meiosis. They are inherited clonally and do not obey Mendelian law (Jones and Houben, 2003). Two key factors are involve in maintenance of Bs, its transmission rate (i.e., drive) and effects on fitness. It is believed unlikely that young extra chromosomes lacking drive or beneficial effects (even being neutral) might invade a population and become B chromosomes (Camacho et al. 1997; Camacho, 2005).

Today more than 15% of eukaryotic species including 500 animal species have been reported to posses B chromosomes (Bs) (Camacho, 2005). Within the same population not all individuals carry B chromosomes and their number can differ between individuals (e.g. *Vulpes vulpes*, 2n = 34 As + 0–8 Bs; *Rattus rattus*, 2n = 42 + 0–5 Bs). Some species have different morphological types of Bs exist within a single species . They can surprisingly exceed the number of As in some species (e.g. Zea mays, 2n = 20 As + 0–34 Bs) (Liehr et al. 2008).

In most species which carry Bs, the mitotic transmission of Bs during growth and development is normal and hence all cells carry the same number of Bs within the individual. However, several exceptional studies stated that some Bs are mitotically unstable and can vary in numbers in specific tissues and/or organs. For example, in the grasses (*Aegilops speltoides* and *A. mutica*), Bs exist in aerial organs but not in roots (Mendelson and Zohary, 1972; Ohta, 1996). Bs can also be distributed differently between genders from species to species. Normally Bs exist in both gender, but sometime the frequency of Bs are higher in one sex. In some species the Bs are present either in males only (eg. *Moenkhausia sanctaefilomenae)* (Portela-Castro et al. 2000) or females only (eg. *Astyanax scabripinnis paranae)* (Maistro et al. 1992; Mizoguchi and Martins-Santos, 1997).

B chromosomes are composed of repetitive DNAs such as rRNA genes (Houben et al. 2005; Ruiz-Estévez et al. 2012), tandemly arranged repetitive elements (Potapov et al. 1990), LINEs (long interspersed nuclear elements), SINEs (short interspersed nuclear elements) (Peppers et al. 1997), interstitial telomeric sequences (Wurster et al. 1988; Szczerbal et al. 2003). The heterochromatic

nature of B chromosomes gives the idea that these elements were mediated genetically inert and their presence were not needed for survival or reproduction of the individuals (Camacho, 2005). But recently comparative sequence analysis of the A and B chromosomes of rye plant (*Secale cereale*) and cichid fish *(Astatotilapia latifaciat*a) claim the inert nature of supernumerary chromosomes. These studies concluded that B chromosome has gained a diverse range of repeat sequences and protein-coding genes (Martis et al. 2012; Valente et al. 2014).

Different studies have revealed that B chromosomes keep transcriptionally active DNA sequences that could play some role in variety of functions, such as the discovery of proto-oncogenes and tumor-suppressor genes in the B chromosomes of canid species(Graphodatsky et al. 2005; Makunin et al. 2014), H3 and H4 histone genes in those of the migratory locust (Teruel et al. 2010), other protein-coding genes in the B chromosomes of a cichlid fish and the Siberian roe deer (*Capreolus pygargus)* (Trifonov et al. 2013; Yoshida et al. 2011). In addition, Valente *et al*. investigated the gene content of B chromosomes in cichlid fish (*Astatotilapia latifasciata)* accomplice with different functions.

The numerical frequency of Bs carrying individuals shows phenotypic effects in some species (Jones, 1982; Green, 1990). The small number of the B chromosomes seems to have no impact on the phenotype while in a high number they can effect the phenotype (Bosemark, 1957b; Gonzalez-Sanchez. et al. 2004). B chromosomes are linked with both negative and positive effects. In grasshopper (*Myrmeleotettix maculatus)*, they likely prevent animal development (Harvey and Hewitt, 1979) and sperm dysfunction (Hewitt et al. 1987). A positive behavior was also found in some organisms for example as in rice (*Oryza sativa*), Bs play role on plant height, weight of grain, and length of its panicle (Cheng et al. 2000). The presence of B chromosomes in maize (*Zea mays L*), alters the recombination frequency of A chromosomes (Rhoades, 1968).

The B chromosome originated as a by-product of A chromosome evolution of either the same or related species (Camacho et al. 2000). The study on Bs origin of *Canidae* identified that it carry several chromosomal regions of domestic dog that show co-hybridization to wild canid B chromosomes (Becker et al. 2015). Recently, the rise of next generation sequencing confessed that the B chromosomes of fish species *Astatolilapia latifasciata* and *Astyanax paranae* were evolved from multiple As (Silva et al. 2014; Valente et al. 2014). Sequencing of rye B chromosome showed that the B chromosome was originated from A chromosomes 3R and 7R(Martis et al. 2012). Kao et al. (2015) using the Random Amplified Polymorphic DNA (RAPD) technology, revealed that four short repetitive sequences were found to locate on both A and B chromosomes.

Acknowledgement

This work was supported by grants from Sao Paulo Research Foundation (FAPESP: process number: 2014/17683-6).

Assembly and Data files

Assembly and sequencing data of PCR products will be made available in NCBI data base soon after the acceptance of paper.

Supplementary Files

Additional files to support our results are organised as the attached PDF file.

References

(include in thesis, will be organised according to Journal format)

7. Thesis conclusion

- The Illumina reads coverage obtained for several samples (including males and females with B+ and B- genotypes) will be useful to generate a population genomics scenario for the species, including the B chromosome polymorphism.
- The development of the present research provides the first *de novo*, both B+ and B- genome, assemblies of *A. latifasciata.* These genomes are to be explored as a reference for future analysis involving evolutionary and applied genomics. Both genomes also open the perspective to advance understanding the role of B chromosomes in evolution of African cichlids
- We have identified nucleotide polymorphism to search for genome variations among B+ and B- individuals. We observed a high scale genomic diversity in all analyzed genomes showing a high rate of population polymorphism. This analysis suggests that the males were comparatively under lower selective pressure than female with no evident effect of B sequences.
- We also identified that duplication events generate a higher number of copies for the Indian Hedgehog b (*Ihhb*) gene and 45S rRNA gene-cluster in B+ genome. Single nucleotide polymorphism for the corresponding genes were detected in B+ sequencing data. The B chromosome does not influence the frequency of SNPs in the genome, however the presence of B specific polymorphisms in duplicated B copies of *Ihhb* gene are importenat in its duplication.
- Structural variations associated with B chromosomes were reported which provoke a hypothesis that supernumerary arise due to the series of these polymorphic events.

10. References

Alekseyev MA (2008). Multi-break rearrangements and breakpoint re-uses: from circular to linear genomes. Journal of Computational Biology. 15:1117–1131.

Alekseyev MA and Pevzner PA (2008). Multi-break rearrangements and chromosomal evolution. Theoretical Computer Science. 395: 193–202.

Alfenito MR and Birchler JA (1993). Molecular characterization of a maize B chromosome centric sequence. Genetics. 135: 589–597.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol. 215: 403–410.

Amos A and Dove G (1981). The distribution of repetitive DNAs between regular and supernumerary chromosomes in species of Glossina (tsetse): a two-step process in the origin of supernumeraries. Chromosoma. 81:673-690.

Bakkali M, Camacho J P M (2004). The B chromosome polymorphism of the grasshopper Eyprepocnemis plorans in North Africa: III. Mutation rate of B chromosomes. Heredity. 92:428-433.

Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A. (2013). Formation and Expression of Pseudogenes on the B Chromosome of Rye. *The Plant Cell*. 25:2536-2544.

Barnes MR, Gray IC (2003). Bioinformatics for geneticists. United Kingdom: Wiley.

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.

Bentley DR (2000). The Human Genome Project - an overview. Med Res Rev. 20: 189-96.

Bickhart DM, Liu GE (2014). The challenges and importance of structural variation detection in livestock. Front Genet. 5:37

Boetzer M and Pirovano W (2012). Toward almost closed genomes with GapFiller. Genome Biology. 6:R56.

Bosemark NO (1957b). On accessory chromosomes in Festuca pratensis 5. Influence of accessory chromosomes on fertility and vegetative development. Hereditas. 43:211-235.

Botstein D, White RL, Skolnick M, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 32 (3):314– 331.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I et al (2013). Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. GigaScience. 2:10.

Branton D, Deamer DW, Marziali A, Bayley H, Benner SA et al (2008). The potential and challenges of nanopore sequencing. Nature Biotechnology. 26:1146–1153.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S et al (2014). The genomic substrate for adaptive radiation in African cichlid fish. Nature. 513:375-381.

Burt A and Trivers RL (2006). Genes in Conflict : The Biology of Selfish Genetic Elements. Belknap Press, Harvard.

Cabral-de-Mello DC, Valente GT, Nakajima RT, Martins C (2012). Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. Chromosome Research. 20: 279-292.

Cabrero J, López-León M, Ruíz-Estévez M, Gómez R, Petitpierre E, Rufas J, and Massa B (2014). B1Was the Ancestor B Chromosome Variant in the Western Mediterranean Area in the GrasshopperEyprepocnemis plorans. Cytogenetic and Genome Research. 142: 54-58.

Cabrero, J. Lo´pez–Leo´n MD, Go´mez R et al (1997). Geographical distribution of B chromosomes in the grasshopper Eyprepocnemis plorans, along a river basin, is mainly shaped by non-selective historical events. Chromosome Res. 5:194–198.

Camacho JP, Sharbel T, Beukeboom L (2000). B-chromosome evolution. Phil Trans R Soc Lond B. 355:163–178.

Camacho JP, Shaw MW, López-León MD, Pardo MC, Cabrero J (1997). Population dynamics of a selfish B chromosome neutralized by the standard genome in the grasshopper Eyprepocnemis plorans. The American Naturalist. 149:1030–1050.

Camacho JPM (2005). B Chromosomes. In: Gregory TR, editor. The evolution of the genome. Amsterdam: Elsevier Academic Press. 223–286.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research., 18(1):, 188–196.

Cargill M, Altshuler D, Ireland J, Irland J, Sklar J, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet. 22:231 – 238.

Carroll SB (1995). Homeotic genes and the evolution of chordates. Nature (London). 376:479–485.

Carvalho AB, Vicoso B, Russo CAM, Swenor B, Clark AG (2015). Birth of a new gene on the Y chromosome of Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America. 112:12450–12455.

Cavalli-Sforza LL, Bodmer WF (1971). The Genetics of Human Populations. San Francisco: W.H. Freeman and Company.

Chen W, Ullmann R, Langnick C, Menzel C, Wotschofsky Z, Hu H, Döring A, Hu Y, Kang H, Tzschach A, Hoeltzenbein M, Neitzel H, Markus S, Wiedersberg E, Kistner G, van Ravenswaaij-Arts CMA, Kleefstra T, Kalscheuer VM, Ropers HH (2010). Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. Eur J Hum Genet. 18:539–43.

Cheng ZK, Yu HX, Yan HH, Gu MH, and Zhu LH (2000). B chromosome in a rice aneuploid variation. Theoretical and Applied Genetics. 101: 564-568.

Cheong WH, Tan YC, Yap SJ, Ng K (2015). ClicO FS: an interactive web-based service of Circos. Bioinformatics. 31:3685-3687.

Chiavarino AM, Rosato M, Naranjo CA, Camara Hernandez H and Poggio L (1995). B chromosome polymorphism in N. Argentine populations of maize. Maize Genet Coop News Lett. 69: 94.

Chung A, Kim S, Kim E, Kim D, Jeong I, cha YR, Bae YK, park SW, Lee J, Park HC (2013). Indian hedgehog B function is required for the specification of oligodendrocyte progenitor cells in the zebrafish CNS. J Neurosci. 33:1728-1733.

Danecek P, Auton A, Abecasis G, Albers A.C, Bank E, Depristo M.A, Handsaker R.E, Lunter G (2011). The variant call format and VCFtools. Bioinformatics. 27:2156-2158.

Datta AK, Mandal A, Das D, Gupta A, Saha R, Paul S (2016). B chromosomes in angiosperm—a review. Cytol Genet. 50: 60.

Eickbush DG, Eickbush TH, Werren JH (1992). Molecular characterization of repetitive DNA sequences from a B chromosome. Chromosoma. 101:575-583.

Ekker SC, Ungar AR, Greenstein P, von Kessler DP, Porter JA, Moon RT, Beachy PA (1995).Patterning activities of vertebrate hedgehog proteins in the developing eye and brain. Curr Biol. 5:944–955.

Ellegren H (2014). Genome sequencing and population genomics in non model organisms. Trends in Ecology & Evolution. 29:51–63.

Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M (2008). Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster. Science. 320:1629–163.

Fan JB, Chee MS, and Gunderson KL (2006). Highly parallel genomic assays. Nature Review Genetics. 7:632–644.

Fan S, Meyer A (2014). Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Front Genet*. 5:163.

Fantinatti BEA and Martins C (2016). Development of chromosomal markers based on next-generation sequencing: the B chromosome of the cichlid fish Astatotilapia latifasciata as a model. BMC Genetics. 17:119.

Fantinatti BEA, Mazzuchelli J, Valente GT, Cabral-de-Mello DC, Martins C (2011). Genomic content and new insights on the origin of the B chromosome of the cichlid fish Astatotilapia latifasciata. Genetica. 139:1273-1282.

Feldberg E, Bertollo LAC (1984). Discordance in chromosome number among somatic and gonadal tissue cells of Gymnogeonhagus balzani. Brazilian Journal of Genetics. 4:639-645.

Feldberg E, Porto JIR, Brinn MNA, Mendonça MNC, Benzaquem DC (2004). B- Chromosomes in Amazonian cichlid species. Cytogenetic and Genome Research. 106:195-198.

Franks K, Houben A, Leach CR, Timmis JN (1996). The molecular organization of a B chromosome tandem repeat sequence from Brachycome dichromosomatica. Chromosoma. 105:223-230.

Friebe B, Jiang J, Gill B (1995). Detection of 5S rDNA and other repeated DNA on supernumerary B chromosomes of Triticum species (Poaceae). Pl Syst Evol. 196:131–139.

Genner MJ, Turner GF (2005). The mbuna cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. Fish Fish. 6:1-34.

Gonzalez-Sanchez M, Chiavarino M, Jimenez G, Manzanero S, Rosato M, Puertas MJ (2004). The parasitic effects of rye B chromosomes might be beneficial in the long term. Cytogenetic and genome research. 106:386-393.

Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK et al (2011). Finished genome of the fungal wheat pathogen Mycosphaerella graminicola reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genetics. 7: e1002070.

Graphodatsky AS, Kukekova AV, Yudkin DV, Trifonov VT, Vorobieva NV, Beklemisheva VR, Perelman PL, Graphodatskaya DA, Trut LN, Yang F, Ferguson-Smith MA, Acland GM and Aguirre GD (2005). The proto-oncogene C-KIT maps to canid B-chromosomes. Chromosome Res. 13:113-12.

Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod DL, Chevraud JM, Ley TJ (2007). A high-resolution map of segmental DNA copy number variation in the mouse genome. PLoS Genet. 3(1):e3.

Green DM (1990). Muller Ratchet and the evolution of supernumerary chromosomes. Genome. 33:818-824.

Gudbjartsson DF, Sulem P, Helgason H, Gylfason A, Gudjonsson SA , Zink F , Oddson A, Magnusson G, Halldorsson BV, Hjartarson E, Sigurdsson GT1, Kong A, Helgason, Masson G, Magnusson OT, Thorsteinsdottir, Stefansson K (2015). Sequence variants from whole genome sequencing a large group of Icelanders. Scientific Data. 2:150011.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013). QUAST: quality assessment tool for genome assemblies.Bioinformatics. 29:1072-1075.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet. 22 (3):239 – 247.

Harvey AW and Hewitt GM (1979). B chromosomes slow development in a grasshopper. Heredity. 42:397-401.

Henriques-Gil N and Arana P (1990). Origin And Substitution Of B Chromosomes In The Grasshopper Eyprepocnemis Plorans. Evolution. 44:747-753.

Henson J, Tischler G, Ning Z (2012). Next-generation sequencing and large genome assemblies. Pharmacogenomics. 13(8):901–915.

Hewitt GM, East TM, Shaw MW (1987). Sperm dysfunction produced by B chromosomes in the grasshopper Myrmeleotettix maculatus. Heredity. 58:59-68.

Holland PWH (1994). Homeobox genes in vertebrate evolution . BioEssays .14:267–273.

Houben A, Banaei-Moghaddam AM, Klemme S, Timmis JN (2013). Evolution and biology of supernumerary B chromosomes. Cell Mol Life Sci. 71:467-478.

Houben A, Verlin D, Leach CR, Timmis JN (2001). The genomic complexity of micro B chromosomes of Brachycome dichromosomatica. Chromosoma. 110: 451–459.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M et al (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature. 496:498-903.

Ijiri S, Kaneko H, Kobayashi T, Wang DD, Sakai F, Paul-Prasanth B, Nakamura M, Nagahama Y (2007). Sexual Dimorphic Expression of Gene in Gonads during early differentiation of a Teleost Fish, the Nile Tilapia Oreochromis niloticus. Bio Reprod. 78:333–341.

Ingham PW, McMahon AP (2001). Hedgehog signaling in animal development: paradigms and principles. Genes Dev. 15: 3059–3087.

Iskow RC, Gokcumen O, Lee C (2012). Exploring the role of copy number variants in human adaptation. *Trends Genet*. 28, 245–257.

Jones JDG and Flavell RB (1982). The structure, amount and chromosomal localization of defined repeated DNA sequences in species of the genus Secale. Chromosoma. 86:613-641.

Jones N and Houben A (2003). B chromosomes in plants: escapees from the A chromosome genome? Trends in Plant Science. 8:417-423.

Jones NR, Gas WD, Ben AH (2008). A Century of B Chromosomes in Plants: So What?. Annals of Botany. 101: 767 –775, 2008.

Jones RN and Rees H (1982). B chromosomes. (London: Academic Press).

Kao KW, Lin CY, Peng SF, Cheng YM (2005). Characterization of four B-chromosome-specific RAPDs and the development of SCAR markers on the maize B-chromosome. Mol Genet Genomics. 2015;290(2):431–41.

Kasahara M, Naruse K, Sasaki S, Nakatan Yi, Qu W , Ahsan B et al (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature. 447:714-719.

Keane TM, Wong K, Adams DJ, Flint J, Reymond A, Yalcin B (2014). Identification of structural variation in mouse genomes. Front Genet. 5:192.

Kellis M, Birren BW, Lander ES (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature. 428:617–24.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P et al (2008). Mapping and sequencing of structural variation from eight human genomes. Nature. 453:56–64.

Kingsford C, Schatz M, Pop M (2010). Assembly complexity of prokaryotic genomes using short reads. BMC Bioinformatics. 11:21.

Klemme S, Banaei-Moghaddam AM, Macas J, Wicker T, Novák P, Houben A (2013). High-copy sequences reveal distinct evolution of the rye B-chromosome. New Phytol. 199:550–558.

Konerat JT, Thums J, Vanessa B, Lucas B, Martins-Santos IC, Margarido VP (2014). B chromosome and NORs polymorphism in Callichthys callichthys (Linnaeus, 1758) (Siluriformes: Callichthyidae) from upper Paraná River, Brazil. Neotropical Ichthyology. 12:603-609.

Korbel JO, Urban AE, Affourtit JP, et al (2007). Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science (New York, NY)*. 318:420-426.

Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009). Circos: An information aesthetic for comparative genomics. Genome Res. doi: 10.1101/gr.092759.109.

Kumar S, Balczare KA, Zhi-Chun L (1996). Phylogenetic analyses of Sonic hedgehog (Shh) and Hoxd-10 genes from 18 cyprinid fish species closely related to the zebrafish. Genetics 142:965–972.

Kumar S, Balczarek K, Lai Z (1996). Evolution of the hedgehog gene family. Genetics. 142:965–1037.

Kuroiwa A, Terai Y, Kobayashi N, Yoshida K, Suzuki M, Nakanishi A, Matsuda Y, Watanabe M, Okada N (2014). Construction of chromosome markers from the Lake Victoria cichlid Paralabidochromis chilotes and their application to comparative mapping. Cytogenet Genome Research. 142:112–120

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357-359.

Leach CR, Houben A, Bruce F, Pistrick K, Demidov D, Timmis JN (2005). Molecular evidence for transcription of genes on a B chromosome in Crepis capillaris. Genetics 171: 269–278

Liehr T, Mrasek K, Kosyakova N, Ogilvie CM, Vermeesch J, Trifonov V, Rubtsov N (2008). Small supernumerary marker chromosomes (sSMC) in humans; are there B chromosomes hidden among them. Mol Cytogenet. 1:12.

Lin G, Chai J, Yuan S, Mai C, Cai L, Murphy R.W, Zhou W, Luo J (2016). VennPainter: A Tool for the Comparison and Identification of Candidate Genes Based on Venn Diagrams. PLoS One. 11: e0154315.

Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin H, Pong R, Lin D, Lu L, Law M (2012). Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology. 2012:251364.

Loh YHE, Katz LS, Mims MC, Kocher TD, Yi SV, Streelman JT (2008). Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. Genome Biology, 9:R113.

López-León M, Cabrero J., Dzyubenko, V, Bugrov A, and Karamysheva T (2008). Differences in ribosomal DNA distribution on A and B chromosomes between eastern and western populations of the grasshopper Eyprepocnemis plorans plorans. Cytogenetic and Genome Research. 121:260-265.

López-León MD, Cabrero J, Pardo MC, Viseras E, Camacho JP, Santos JL (1993). Generating high variability of B chromosomes in Eyprepocnemis plorans (grasshopper). Heredity. 71:352-362.

Lukashin AV and Borodovsky M (1998). GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26:1107–1115.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J et al (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 1:18.

Mahul C, James GBB, Anthony DL, Emerson JJ (2015). A practical guide to de novo genome assembly using long reads. BioRxiv.

Maistro EL, Foresti F, Oliveira C, Toledo LFD (1992). Occurrence of macro-B chromosomes in Astyanax scabripiMartins-Santos IC, Portela-Castro ALB, Julio HF Jr (1995). Chromosome analysis of 5 species of the Cichlidae family (Pisces, Perciformes) from the Paraná River. Cytologia. 60:223–231.

Makunin AI, Dementyeva PW , Graphodatsky AS, Volobouev VT , Kukekova AV, Trifonov VA (2014). Genes on B chromosomes of vertebrates. Molecular Cytogenetics. 7:99.

Marques DF, Conte MA, Fantinatti BEA, Valente GT, Nakajima RT, Coan RLB, Kocher TD, Martins C. B chromosomes effects on gene transcription in the cichlid fish Astatotilapia latifasciata (unpublihed).

Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutzer T, Scholz U, Gundlach H, Wicker T, Simkova H, Novak P, Neumann P, Kubalakova M, Bauer E, Haseneyer G, Fuchs J, Dolezel J, Stein N, Mayer KF, Houben A (2012). Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. Proc Natl Acad Sci USA. 109:13343–13346.

Mathieu Blanchette (2001). Evolutionary Puzzles: An Introduction to Genome Rearrangement. Computational Science – ICCS. 2074:1003-1011.

Matzke M, Varga F, Berger H, Schernthaner J, Schweizer D, Mayr B, Matzke AJM (1990). A tandemlyrepeated sequence isolated from nuclear envelopes of chickenerythrocytes is located predominantly on microchromosomes. Chromosoma. 99: 131-137

Mazzuchelli J, yang F, Kocher TD, Martins C (2011). Comparative cytogenetic mapping of *Sox*2 and *Sox14* in cichlid fishes and inferences on the genomic organization of both genes in vertebrates. Springer Link. 19: 657–667.

McClintock B, T. Kato TA, Blumenschein A (1981). Chromosome Constitution of the Races of Maize. Colegio de Postgraduados, Chapingo, México.

Mendelson D and Zohary D (1972). Behavior and transmission of supernumerary chromosomes in Aegilops speltoides. Heredity. 29:329–339.

Metzker ML (2005). Emerging technologies in DNA sequencing. Genome Research. 15:1767-1776.

Meyer A, Van de Peer Y (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays. 27:937–945.

Mizoguchi SMHN and Martins-Santos IC (1997). Macro- and microchromosomes B in females of Astyanax scabripinnis (Pisces, Characidae). Hereditas. 127:249-253.

Muñoz-Pajares AJ, Martínez-Rodríguez L, Teruel M, Cabrero J, Camacho J PM, Perfectti F (2011). A Single, Recent Origin of the Accessory B Chromosome of the Grasshopper Eyprepocnemis plorans. Genetics. 187:853–863.

Naiara RE, Michael H, Ana MA (2012). Bioinformatics for High Throughput Sequencing. Springer. DOI 10.1007/978-1-4614-0782-9.

Naranjo CA, Chiavarino AM, Rosato M, Quintela Fernández E, Poggio L (1995). Tamaño del genoma y polimorfismo para cromosomas B en razas nativas argentinas y bolivianas de maiz. 969–980.

Noor MA, Grams KL, Bertucci LA, Reiland J (2001). Chromosomal inversions and the reproductive isolation of species. Proc Natl Acad Sci U.S.A. 98:12084–12088.

O'Quin CT, Drilea AC, Conte MA, Kocher TD. Mapping of pigmentation QTL on an anchored genome assembly of the cichlid fish, Metriaclima zebra. BMC Genomics.14: 287.

Ohta S (1996). Mechanisms of B-chromosome accumulation in Aegilops mutica Boiss. Genes Genet Syst. 71(1):23–29.

Oliveira C, Saboya SMR, Foresti F, Senhorini JA, Bernardino G (1997). Increased B chromosome frequency and absence of drive in the fish Prochilodus lineatus. Heredity. 79:473-476.

Parra G, Bradnam K, Korf I (2007). CEGMA: a pipeline to accurately annotate core genes eukaryotic genomes. Bioinformatics. 23 :1061-1067.

Pauls E, Bertollo LAC (1983). Evidence for a system of supernumerary chromosome in Prochilodus scrofa (Pisces, Prochilodontidae). Caryologia. 36:307-314.

Pellegrino KCM, Rodrigues MC, Yonenaga-Yassuda Y (1999). Chromosomal polymorphisms due to supernumerary chromosomes and pericentric inversions in the eyelidless microteiid lizard Nothobachia ablephara. Chromosome Research.7:247-254.

Peng SF, Lin YP, Lin BY (2005). Characterization of AFLP sequences from regions of maize B chromosome defined by 12 B-10L translocations. Genetics. 169:375-388.

Peppers JA, Wiggins LE, Baker RJ (1997). Nature of B chromosomes in the harvest mouse Reithrodontomys megalotis by fluorescence in situ hybridization (FISH). Chromosome Res. 5:475–479.

Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009). Metagenomic pyrosequencing and microbial identification. Clinical Chemistry. 55:856–866.

Petrov DA (2002). Mutational equilibrium model of genome size evolution. Theor Popul Biol. 61:533–46.

Pinkel D, Landegent J, Collins C, et al (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proceedings of the National Academy of Sciences of the United States of America. 85: 9138-9142.

Poletto AB, Ferreira IA, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, Ribeiro HB, Venere PC, Nirchio M, Kocher TD, Martins C (2010). Chromosome differentiation patterns during cichlid fish evolution. BMC Genet. 11: 50.Poletto AB, Ferreira IA, Martins C (2010a). The B chromosome of the cichlid fish Haplochromi sobliquidens harbors 18S rRNA genes. BMC Genetics. 11:1.

Pop M and Salzberg SL (2008). Bioinformatics challenges of new sequencing technology. Trends Genetics. 24:142–149.

Portela-Castro ALD, Juilo HF, Nishiyama PB (2000). New occurrence of microchromosomes B in Moenkhausia sanctaefilomenae (Pisces, Characidae) from the Parana River of Brazil: analysis of the synaptonemal complex. Genetica 110:277-28.

Potapov VA, Solov'ev VV, Romashchenko AG, Sosnovtsev SV, Ivanov SV (1990). Features of the structure and evolution of complex, tandemly organized Bsp-repeats in the fox genome. I. Structure and internal organization of the BamHI-dimer. Mol Biol. 24:1649–1665.

Qi J, Zhao F (2011). inGAP-sv: a novel scheme to identify and visualize structural variation from paired end mapping data. Nucleic Acids Res. 39(Web Server issue):W567-575.

Quinlan AR and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841-842.

R. Chikhi, P. Medvedev, Informed and automated *k*-mer size selection for genome assembly, *Bioinformatics* (2014) 30 (1): 31-37.

Ramirez J, dundar F, Diehl s, Gruning B, Manke T (2014). deepTools: a flexible platform for exploring deep-sequencing data. doi: 10.1093/nar/gku365.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 28:i333-i339.

Rhoades M (1968). Studies on the cytological basis of crossing over. In: Peacock W, Brock P, editors. Replication and Recombination of Genetic Material. Canberra: Australian Acad Sci. 229–41.

Riva A and Kohane IS (2002). SNPper: retrieval and analysis of human SNPs. Bioinformatics. 18:1681-1685.

Robert E and Jochen BWW (2014). A field guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications. 7:1026–1042.

Rosato M, Chiavarino AM, NaranjoCA, Cámara Hernández J, Poggio (1998). Genome size and numerical polymorphism for the B chromosome in races of maize (Zea mays ssp. mays, Poaceae). Am J Bot. 85: 168–174.

Rubtsov NB, Karamysheva TV, Andreenkova OV, Bochkaerev MN, Kartavtseva IV, Roslik GV, Borissov YM (2004). Comparative analysis of micro and macro B chromosomes in the Korean field mouse Apodemus peninsulae Rodentia, Murinae) performed by chromosome microdissection and FISH Cytogenet. Genome Res 2004. 106:289–294.

Ruíz-Estévez M, López-León MD, Cabrero J, Camacho JPM (2012) B-Chromosome Ribosomal DNA Is Functional in the Grasshopper Eyprepocnemis plorans. PLoS ONE. 7:e36600.

Sandery MJ, Forster JW, Blunden R, Jones RN (1990). Identi¢cation of a family of repeated sequences on the rye B chromosome. Genome. 33: 908-913.

Sanger F; Nicklen S; Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci. 74: 5463–5467.

Schartl M, Walter RB, Shen Y, Garcia T, Catchen J et al (2013). The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. Nat Genet. 45:567–572.

Schatz MC, Delcher AL, Salzberg SL (2010). Assembly of large genomes using second-generation sequencing. Genome. 9:1165-1173.

Schmid M, Ziegler CG, Steinlein C, Nanda I, Haaf, T (2002). Chromosome banding in amphibia - XXIV. The B chromosomes of Gastrotheca espeletia (Anura, Hylidae). Cytogenetic and genome research. 97:205-21.

Schork NJ, Fallin D, Lanchbury S (2000). Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet .58:250 – 264.

Seo JH, Bae GH, Park DH, Kim BS, Lee JW, Lee JI, Kim KH, Lee SK, Seo BB (2013). Sequence polymorphisms in ribosomal RNA genes and variations in chromosomal loci of Oenothera odorata and O. laciniata. Genes Genom. 35:117–124.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE (1997). Ethnic affiliation estimation by use of population-specific DNA markers. Am J Hum Genet. 60:957 – 964.

Silva DM, Pansonato-Alves JC, Utsunomia R, Araya-Jaime C, Ruiz-Ruano FJ, Daniel SN, Hashimoto DT, Oliveira C, Camacho JPM, Porto-Foresti F, Foresti F (2014). Delimiting the Origin of a B Chromosome by FISH Mapping, Chromosome Painting and DNA Sequence Analysis in (Teleostei, Characiformes). PloS One. 9:e94896.

Silva MJJ, Yonenaga-Yassuda Y (1998). Heterogeneity and meiotic behaviour of B and sex chromosomes, banding patterns and localization of (TTAGGG)n sequences by FISH, in the Neotropical water rat Nectomys (Rodentia, Cricetidae). Chromosome Res. 6: 455-462.

Simpson JT and Pop M (2015). The Theory and Practice of Genome Sequence Assembly. Annu Rev Genomics Hum Genet. 16:153-72.

Sims DI, Sudbery N E, IlottA, Heger CP, Ponting (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nature Reviews Genetics. 15:121–132

Smit AF (1996). The origin of interspersed repeats in the human genome. Curr Opin Genet Dev. 6:743–748.

Soon WW, Hariharan M and Michael P Snyder (2013). High-throughput sequencing for biology and medicine. Molecular System Biology. 9:640.

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

Stiassny MLJ (1991). Phylogenetic intrarelationships of the family Cichlidae: an overview. Cichlid fishes. Behaviour, ecology and evolution. London: 1-35.

Stitou S, de La Guardia RD, Jiménez R, Burgos M (2000). Inactive ribosomal cistrons are spread throughout the B chromosomes of Rattus rattus (Rodentia, Muridae). Implications for their origin and evolution. Chromosome Res. 8:305–311.

Strittmatter WJ and Roses AD (1996). Apolipoprotein E and Alzheimer's disease. Annu Rev Neurosci. 19:53-77.

Szczerbal I and Switonski M (2003). B chromosomes of the Chinese raccoon dog (Nyctereutes procyonoides procyonoides Gray) contain inactive NOR-like sequences. Caryologia. 56:213–216.

Teruel M, Cabrero J, Perfectti F, Camacho JPM (2010). B chromosome ancestry revealed by histone genes in the migratory locust. Chromosoma. 119:217–225.

Treangen TJ and Salzberg SL (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews. Genetics, 13:36–46.

Trifonov VA, Dementyeva PV, Larkin DM, O'Brien PC, Perelman PL, Yang F, Ferguson-Smith M, Graphodatsky AS (2013). Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (Capreolus pygargus). BMC Biolology. 6:90.

Valente GT, Conte MA, Fantinatti BEA, Cabral-de-melo DC, Carvalho R, Vicari MR, Kocher TD, Martins C (2014). Origin and Evolution of B Chromosomes in the Cichlid Fish Astatotilapia latifasciata Based on Integrated Genomic Analyses. Molecular Biology Evolution. 31:2061-72.

Vicente VE, Moreira-Filho O, Camacho JP (1996). Sex-ratio distortion associated with the presence of a B chromosome in Astyanax scabripinnis (Teleostei, Characidae). Cytogenetics and cell genetics. 74:70-75.

Wang Z, Gerstein M, Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 10:57–63.

Wheat CW (2010). Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica. 138:433–451.

Wilkes TM, Francki MG, Langidge P, Karp A, Jones RN, Forster JW (1995). Analysis of rye B-chromosome structure using fluorescence in situ hybridization (FISH). Chromosome Res. 3:466 -472.

Wilson EB (1907a). The supernumerary chromosomes of Hemiptera. Science. 26:870–871.

Wold B and Myers RM (2008). Sequence census methods for functional genomics. Nature Methods. 5:19–21.

Wurster-Hill DH, Ward OG, Davis BH, Park JP, Moyzis RK, Meyne J (1988). Fragile sites, telomeric DNA sequences, B chromosomes, and DNA content in raccoon dogs, Nyctereutes procyonoides, with comparative notes on foxes, coyote, wolf, and raccoon. Cytogenet Genome Res. 49:278–281.

Yandell M and Ence D (2012). A beginner's guide to eukaryotic genome annotation. Nat Rev Genet, 13:329–342.

Yang MQ, Athey BD, Arabnia HR, Sung AH, Liu Q, Yang JY, Deng Y (2009). High-throughput next generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics. BMC Genomics. 10:I1.

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

Zardoya R, Abouheif E, Meyer A (1996). Evolutionary analyses of hedgehog and Hoxd-10 genes in fish species closely related to the zebrafish. Proc Natl Acad Sci. USA. 93:13036–13041.

Zerbino DR and Birney E (2008). Velvet: Algorithms for de novo short read assemblyusing de Bruijn graphs. Genome Research. 18:821–829

Zeyl CW, Green DM (1992). Heteromorphism for a highly repeated sequence in the New Zealand frog Leiopelma hochstetteri. Evolution. 46:1891-1899.

Zurita, S, Cabrero J, Lopez-Leon MD, Camacho JP (1998). Polymorphism Regeneration for a Neutralized Selfish B Chromosome. Evolution. 52: 274.