

RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta tese
será disponibilizado somente a partir
de 23/08/2024.



SÃO PAULO STATE UNIVERSITY “JÚLIO DE MESQUITA FILHO”
BIO SCIENCES INSTITUTE BOTUCATU
POSTGRADUATE PROGRAM IN BIOLOGICAL SCIENCES (GENETICS)

MARYAM JEHANGIR

A NOVEL APPROACH TO ASSEMBLE THE COMPLEX B CHROMOSOME USING A
COMBINATION OF MODERN GENOMICS TECHNOLOGIES

Botucatu, SP
2022

Maryam Jehangir

**A NOVEL APPROACH TO ASSEMBLE THE COMPLEX B CHROMOSOME USING A
COMBINATION OF MODERN GENOMICS TECHNOLOGIES**

Thesis submitted to the Postgraduate Program in
Biological Sciences (Genetics) of the Institute of Bio
sciences of Botucatu from the State University of São
Paulo "Júlio de Mesquita Filho" for achievement of the
Ph.D. in Biological Sciences (Genetics).

Advisor: Prof. Dr. Cesar Martins

Botucatu, SP
2022

Acknowledgment

I acknowledge all, including the institutions, funding agencies, mentors, committee members, colleagues and my family who in one way or another contributed in the completion of this PhD thesis.

I gratefully acknowledge the Post-Graduate Program in Biological Sciences (Genetics) of the University São Paulo State University, for the opportunity to expand my knowledge and making it possible for me to accomplish Ph.D. as international student.

This work would not have been possible without the financial support of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The project of my PhD was funded by CAPES (process number: 88882.433287/2019.01) for my doctoral scholarship. Additionally, I was awarded a fellowship for research internship abroad (Ph.D. sandwich program, process number: 201042/2020-7). Both CAPES and CNPq are acknowledged for supporting my studies and research.

My special and heartily thanks to my supervisor, Prof. Dr. Cesar Martins who encouraged and directed me in this work and significantly contributed to my academic training. During my academic tenure of Ph.D., Prof. Cesar 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. His challenges brought this work towards a completion. It is with his supervision that this work was accomplished. For any faults I take full responsibility.

I am also deeply thankful to Assoc. Prof. Dr. Kornorn Srikulnath who accepted me in his laboratory for the accomplishment of my research internship abroad, allowing me to experience advanced training in an international environment. I also acknowledge the Department of Genetics, Kasetsart University, Thailand for hosting me during the 2 years training, research and academic activities.

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

This thesis was accomplished with the help and support of my fellow lab mates and collaborators, Adauto Lima Cardoso, Jordana Oliveira, Ivan Wolf, Luiz Bovolenta, Camila Moreira and my lab colleagues from Animal Genomics and Bioresource center, Kasetsart University.

I also thank my family who encouraged me and prayed for me throughout the time, and my husband, Farhan Ahmad, for his continued support, encouragement and collaboration in my research. Also not forgetting my daughter, Inshirah, her cute smile gives me more courage.

Maryam Jehangir

Table of Contents

1 INTRODUCTION.....	1
1.1 Genomes and chromosomes: An overview.....	1
1.2 B chromosomes.....	3
1.3 The enigma of B and sex chromosomes.....	5
1.4 Cichlids genomes: Interesting model systems for understanding B chromosomes.....	8
1.5 <i>Astatotilapia latifasciata</i> as an exciting model species for B chromosomics.....	9
1.6 The importance of chromosome-scale genome assembly for investigating B chromosome biology.....	10
1.7 3D genome organization and its importance in investigating the B chromosome biology.....	13
1.8 Rationale of the present study.....	17
2 OBJECTIVES AND HYPOTHESIS.....	19
2.1 General objectives.....	19
2.2 Specific Objectives.....	19
2.3 Hypothesis 1.....	20
2.4 Hypothesis 2.....	20
3 MATERIALS AND METHODS.....	21
3.1 Ethics Statement.....	21
3.2 Fish samples and karyotype.....	21
3.3 PacBio circular consensus library preparation and sequencing.....	21
3.4 Hi-C library preparation.....	22
3.5 Estimation of the genome size and sequencing coverage.....	23
3.6 The draft genome assembly of PacBio reads.....	23
3.7 Polishing draft assembly with Illumina reads.....	24
3.8 Hi-C scaffolding and chromosome-scale assembly.....	25
3.9 Genome annotation.....	27
3.9.1 Repeat element analysis and TEs identification.....	27
3.9.2 Protein-coding gene structural annotation.....	28
3.9.3 Functional annotation.....	28

3.10 B Chromosome identification.....	29
3.10.1 Genome alignments and coverage ratio analysis.....	29
3.10.2 Comparison of Hi-C interactions between 0B and 1B genomes.....	30
3.10.3 Mapping experimentally validated markers on B chromosome assembly.....	30
3.11 3D genome organization and TADs detection.....	31
3.11.1 Comparison of 3D genome structure.....	31
3.11.2 Preprocessing of Hi-C data and TAD prediction.....	31
3.11.3 A/B compartment analysis.....	33
3.11.4 Loop detection.....	33
3.11.5 TADs and Loops size comparisons in both genomes.....	34
3.11.6 Duplication breakpoints at the TADs.....	34
3.11.7 Identification of TADs in representative cichlids species.....	34
3.12 Comparative genomics analysis.....	35
3.12.1 Gene family clustering.....	35
3.12.2 Phylogeny and divergence time.....	35
3.12.3 Gene family evolution.....	36
3.12.4 Genomic structural variation analysis.....	37
3.13 Transcriptomic analysis.....	37
3.13.1 B chromosome genes and differential expression.....	37
3.13.2 Analysis of genes transcription within and outside TADs.....	38
3.14 Mapping of cichlid sex-determination genes in genome assembly.....	39
4 RESULTS.....	40
4.1 PacBio and Hi-C sequencing enabled a reference-quality chromosome-length assembly of <i>A. latifasciata</i> genome.....	40
4.2 Genome assessment confirms the continuity and completeness of assembly.....	45
4.3 Chromosome localization of sex-determining genes suggests polygenic sex determination system in <i>A. latifasciata</i>	48
4.4 Genome annotation and repeatomics revealed a low-level gene density and expansion of LTRs on B.....	49

4.5 An integrative genomic approach allowed the isolation and validation of B chromosome sequences.....	52
4.6 Exploring the 3D genomics identified dynamics of chromatin organization in 1B genome..	55
4.7 Comparative genomics provided adaptive evolutionary insights in cichlids.....	60
4.8 Genomic structural variation analysis revealed a higher rate of duplications in 1B as compared to 0B.....	67
4.9 B chromosome has lower genes expression than A chromosomes.....	69
5 DISCUSSION.....	74
5.1 A chromosome scale cichlid genome for underpinning the B chromosome biology.....	74
5.2 The hidden genomic content on B chromosome displays its evolutionary features.....	75
5.3 Does the B chromosome play any role in reshaping the chromatin conformation inside the cell?.....	80
5.4 Cichlids genomes diversification is indicative of rapid adaptation and speciation.....	84
5.5 The possibility of diverse range of sex determination systems in Lake Victoria cichlids.....	88
6 CONCLUSION.....	90
7 REFERENCES.....	92
8 SUPPLEMENTARY DATA.....	104
8.1 List of supplementary Figures.....	104
8.2 List of supplementary Tables.....	105

LIST OF FIGURES

Figure 1. Karyotype of <i>A. latifasciata</i> with B chromosomes.....	9
Figure 2. Hierarchical genome organization.....	14
Figure 3. HiC as a tool to study the effect of structural variation.....	16
Figure 4. An overview of assembly methods.....	26
Figure 5. Chromosome scale assembly and the 3D model of <i>A. latifasciata</i> genome.....	43
Figure 6. Evaluation of genome assemblies by multiple methods.....	47
Figure 7. Circular visualization of genome assembly.....	51
Figure 8. Validation and confirmation of B chromosome assembly.....	54
Figure 9. The 3D genomic landscape of 0B and 1B genomes.....	57
Figure 10. Evolutionary trajectories of cichlid genomes.....	63
Figure 11. Genomic differences between the 0B and 1B genomes.....	68
Figure 12. The transcriptional profiles of genes.....	73

LIST OF TABLES

Table 1. Summary of obtained sequencing data.....	41
Table 2. A comparison of Illumina, PacBio and Hi-C <i>de novo</i> assemblies statistics.....	46
Table 3. Hi-C interactions of representative cichlids	60
Table 4. Functional enrichment of genes expansion on B chromosome.....	66
Table 5. Chromosome scale comparison of genes expression	71

ABSTRACT

B chromosomes (Bs) are nonvital extra chromosomes found in diverse eukaryotic species including fungi, plants and animals. Among hundreds of investigated species, cichlid genomes offer fascinating models for studying B chromosome biology. Despite the extensive investigations, Bs are poorly understood mainly in relation to mechanisms of their own evolutionary survival as well as their structural and functional impact on genome organization. Our previous studies identified several sequences (genes and repeats) on the B of the cichlid fish *Astatotilapia latifasciata* but the complete genomic map of B chromosome has been missing. Here, we generated a chromosome-scale *A. latifasciata* genome with the B chromosome assembly by adopting integrative approach that combined deep coverage Pacific Biosciences single-molecule real-time (Pacbio long reads), high-throughput chromatin conformation capture (Hi-C) mapping, and Illumina (short-reads) sequencing. The assembled genome spans a total of 0.93 giga base pairs (Gb) genome with contig and scaffold N50 values of 3.4 and 36.2 mega base pairs (Mb) respectively. Compared with our previous Illumina based assembly, this upgraded genome is much more complete, and accurate. The annotation of core eukaryotic genes and universal single-copy orthologs has also been significantly improved and a total 150 Mb region has now been recovered, which was missing (in the previous assembly). We identified 759 protein-coding genes in the 34 Mb genomic content of B chromosome, of which at most of the genes showed a reduced level of expression as compared to A chromosomes. Our results demonstrate a substantial higher amount of transposable elements (TEs) mainly long terminal repeats (LTRs) retrotransposons (mean density 47.6) on the B chromosome as compared to the standard A chromosome set (mean density 10.45). We further applied whole-genome chromosome conformation capture (Hi-C) and in silico modeling methods to characterize the three dimensional (3D) architecture of and A and B chromosomes in the *A. latifasciata* genome. Remarkably, we observed a differential level organization of the chromatin into topologically associated domains (TADs) between 0B and 1B genomes. On a global level, the Hi-C interaction matrices of 1B genome are characterized by a relative gain of long-range and loss of short-range interactions within chromosomes indicating the impact of B chromosome on three-dimensional landscape of genome. Additionally, comparative analysis of *A. latifasciata* genome with other

cichlids genomes enabled us to detect the phylogenomic diversification and chromosomal rearrangements including fusions and inversions providing interesting insights in co-ancestral cichlids gene evolution.

Keywords: Cichlid genome; supernumerary chromosome; Hi-C; 3D genome, Genes evolution

6 CONCLUSION

This study provides advances and contributions towards understanding the genomic composition structure, function and evolution of B chromosome in the cichlid fish. Generating a chromosome scale genome of *A. latifasciata*, our study provides crucial insights into B chromosome biology and reveals the genomic content of B chromosome updating significant additions to the sequences identified previously by Valente et al. (2014). The 3D genomic and bioinformatic approaches employed here broaden the study of karyotypes and chromosomes and fills the knowledge gaps in ways which were not possible using classical or even molecular cytogenetics alone. Long reads sequencing, and chromatin conformation capture technologies create new opportunities for understanding the function and evolution of sequences on B chromosomes. The genome assembly accomplished in the study significantly upgraded the previous version assembly recovering an addition of 150 Mb with a considerable 34 Mb of B chromosome content. The comprehensive genes annotations and repeat annotations using updated models and bioinformatic tools, described herein, will serve as a resource for expanding the compendium of B chromosome and reveal its impact on the host genome. The genes and repeatomic annotation found a high TEs density and low genes density on the B chromosome. Parallel to developing the annotation resource, we provide a methodological framework for assessing the three-dimensional genomic landscape of cichlid genome particularly in the context of the B chromosome dynamics inside the cell. The genomic and transcriptomics scale analyses together with exhaustive assessment of the Hi-C comparisons in the presence of B chromosome, sets the stage for functional studies to disentangle the role of B chromosome and for unlocking the mechanisms essential for driving the chromosome scale evolution. Deeper exploration of Hi-C data also suggests the complexity of genetic mechanisms in organization of TADs and further illustrate the myriad mechanisms by which B chromosome might become major contributor to induce differential TADs interactions in the host genome. Furthermore,

our transcriptomic profiling of B localized genes showed a reduced level expression, indicating the B chromosome may follow the partial inactivation, with exception of certain active genes which can represent advantages for B chromosome survival in the cell.

Finally, the availability of the *A. latifasciata* chromosome scale genome assembly is also of great importance because it is very common fish in the aquarium, which is fortunate because it is now presumed extinct in the wild habitat of African lakes and also occupying a unique evolutionary position among cichlids species. In addition, the high-quality chromosome-level assembly obtained herein provides useful genetic information for accelerating the progress of B chromosome and sex chromosome origin, evolution and composition, as well as functional gene discovery and comparative genomics across the cichlid's family. Despite remaining unanswered questions, our reference genomes are the most complete and highest quality to date for a cichlid species with B chromosome. In broader way, our work demonstrates the need to increase efforts toward achieving chromosome-level assemblies for cichlids and other species to fully understand the complexity and evolution of genomes and chromosomes.

7 REFERENCES

- Ahi EP, Singh P, Duenser A, et al (2019) Divergence in larval jaw gene expression reflects differential trophic adaptation in haplochromine cichlids prior to foraging. *BMC Evol Biol* 19:150. <https://doi.org/10.1186/s12862-019-1483-3>
- Ahmad SF, Jehangir M, Cardoso AL, et al (2020) B chromosomes of multiple species have intense evolutionary dynamics and accumulated genes related to important biological processes. *BMC Genomics* 21:1. <https://doi.org/10.1186/s12864-020-07072-1>
- Ahmad SF, Martins C (2019) The modern view of B chromosomes under the impact of high scale omics analyses. *Cells* 8:156. <https://doi.org/10.3390/cells8020156>
- Akbari OS, Antoshechkin I, Hay BA, Ferree PM (2013) Transcriptome profiling of *Nasonia vitripennis* testis reveals novel transcripts expressed from the selfish B chromosome, paternal sex ratio. *G3 (Bethesda)*. 3:1597–1605. doi: 10.1534/g3.113.007583. PMID: 23893741
- Akdemir KC, Chin L (2015) HiCPlotter integrates genomic data with interaction matrices. *Genome Biol* 16:198. <https://doi.org/10.1186/s13059-015-0767-1>
- Allender CJ, Seehausen O, Knight ME, et al (2003) Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proc. Natl. Acad. Sci. U. S. A.* 100:14074–14079. <https://doi.org/10.1073/pnas.2332665100>
- Arias-Rodriguez L, Ibarra-Castro L, Páramo-Delgadillo S (2008) Los cromosomas mitóticos y meióticos del pez tropical *Petenia splendida* (Cichlidae). *Rev Biol Trop* 56:895–907. <https://doi.org/10.15517/rbt.v56i2.5632>
- Ay F, Noble WS (2015) Analysis methods for studying the 3D architecture of the genome. *Genome Biol* 16:183. <https://doi.org/10.1186/s13059-015-0745-7>
- Babcock M, Pavlicek A, Spiteri E, et al (2003) Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution. *Genome Res* 13:2519–2532. <https://doi.org/10.1101/gr.1549503>
- Bachtrog D, Kirkpatrick M, Mank JE, et al (2011) Are all sex chromosomes created equal? *Trends Genet* 27:350–357. <https://doi.org/10.1016/j.tig.2011.05.005>
- Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A (2013) Formation and expression of pseudogenes on the B chromosome of rye. *Plant Cell* 25:2536–2544. <https://doi.org/10.1105/tpc.113.111856>
- Basso A, Lifschitz E (1995) Size polymorphism of the X-chromosome due to attachment of a B-chromosome in the Medfly, *Ceratitis capitata* (Wied.). *Rev Bras Genet* 18:165–171
- Beladjal L, Vandekerckhove TTM, Muysen B, et al (2002) B-chromosomes and male-biased sex ratio with paternal inheritance in the fairy shrimp *Branchipus schaefferi* (Crustacea, Anostraca). *Heredity (Edinb)* 88:356–360. <https://doi.org/10.1038/sj.hdy.6800061>
- Belton JM, McCord RP, Gibcus JH, et al (2012) Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* 58:268–276. <https://doi.org/10.1016/j.ymeth.2012.05.001>
- Benetta ED, Antoshechkin I, Yang T, et al (2020) Genome elimination mediated by gene expression from a selfish chromosome. *Sci Adv* 6:9808. <https://doi.org/10.1126/sciadv.aaz9808>
- Bentley DR, Balasubramanian S, Swerdlow HP, et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59. <https://doi.org/10.1038/nature07517>
- Bian C, Li J, Lin X, et al (2019) Whole Genome Sequencing of the Blue Tilapia (*Oreochromis aureus*) Provides a Valuable Genetic Resource for Biomedical Research on Tilapias. *Mar Drugs* 17:.. <https://doi.org/10.3390/md17070386>

- Birchler JA, Veitia RA (2012) Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A* 109:14746–14753. <https://doi.org/10.1073/pnas.1207726109>
- Blavet N, Yang H, Su H, et al (2021) Sequence of the supernumerary B chromosome of maize provides insight into its drive mechanism and evolution. *Proc Natl Acad Sci U S A* 118:2104254118. <https://doi.org/10.1073/PNAS.2104254118>
- Boeckmann B, Bairoch A, Apweiler R, et al (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31:365–370. <https://doi.org/10.1093/nar/gkg095>
- Bolzer A, Kreth G, Solovei I, et al (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 3:0826–0842. <https://doi.org/10.1371/journal.pbio.0030157>
- Bonev B, Cavalli G (2016) Organization and function of the 3D genome. *Nat Rev Genet* 17:661–678. <https://doi.org/10.1038/nrg.2016.112>
- Bouwman BAM, de Laat W (2015) Getting the genome in shape: The formation of loops, domains and compartments. *Genome Biol* 16:. <https://doi.org/10.1186/s13059-015-0730-1>
- Boyle S, Gilchrist S, Bridger JM, et al (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* 10:211–219. <https://doi.org/10.1093/hmg/10.3.211>
- Brawand D, Wagner CE, Li YI, et al (2015) The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513:375–381. <https://doi.org/10.1038/nature13726>
- Brockdorff N, Turner BM (2015) Dosage compensation in mammals. *Cold Spring Harb. Perspect. Biol.* 7. doi: 10.1101/cshperspect.a019406. PMID: 25731764; PMCID: PMC4355265.
- Brůna T, Hoff KJ, Lomsadze A, et al (2021) BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma* 3:1–11. <https://doi.org/10.1093/nargab/lqaa108>
- Buchfink B, Reuter K, Drost HG (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>
- Busslinger M, Tarakhovsky A (2014) Epigenetic control of immunity. *Cold Spring Harb. Perspect. Biol.* 6:19307. doi: 10.1101/cshperspect.a019307
- Cabanettes F, Klopp C (2018) D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 2018:. <https://doi.org/10.7717/peerj.4958>
- Camacho C, Coulouris G, Avagyan V, et al (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
- Camacho JPM, Schmid M, Cabrero J (2011) B chromosomes and sex in animals. *Sex Dev* 5:155–166. <https://doi.org/10.1159/000324930>
- Camacho JPM, Sharbel TF, Beukeboom LW (2000) B-chromosome evolution. *Philos Trans R Soc B Biol Sci* 355:163–178. <https://doi.org/10.1098/rstb.2000.0556>
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Cardoso AL, De Almeida Fantinatti BE, Venturelli NB, et al (2019) Epigenetic DNA modifications are correlated with B chromosomes and Sex in the cichlid *astatotilapia latifasciata*. *Front. Genet.* 10
- Cardoso AL, Venturelli NB, da Cruz I, et al (2022) Meiotic behavior, transmission and active genes of B chromosomes in the cichlid *Astatotilapia latifasciata*: new clues about nature, evolution and maintenance of accessory elements. *Mol Genet Genomics*. <https://doi.org/10.1007/s00438-022-01911-4>
- Carlson WR, Roseman RR (1992) A new property of the maize B chromosome. *Genetics* 131:211–223. <https://doi.org/10.1093/genetics/131.1.211>

- Carmello BO, Coan RLB, Cardoso AL, et al (2017) The hnRNP Q-like gene is retroinserted into the B chromosomes of the cichlid fish *Astatotilapia latifasciata*. *Chromosom Res* 25:277–290. <https://doi.org/10.1007/s10577-017-9561-0>
- Carvalho AB (2002) Origin and evolution of the *Drosophila* Y chromosome. *Curr Opin Genet Dev* 12:664–668. [https://doi.org/10.1016/S0959-437X\(02\)00356-8](https://doi.org/10.1016/S0959-437X(02)00356-8)
- Casane D, Rétaux S (2016) Evolutionary Genetics of the Cavefish *Astyanax mexicanus*. *Adv Genet* 95:117–159. <https://doi.org/10.1016/bs.adgen.2016.03.001>
- Chabot D, McKenzie DJ, Craig JF (2016) Metabolic rate in fishes: Definitions, methods and significance for conservation physiology. *J. Fish Biol.* 88:1–9
- Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ (2016) Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res* 44:. <https://doi.org/10.1093/nar/gkw654>
- Chin CS, Peluso P, Sedlazeck FJ, et al (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13:1050–1054. <https://doi.org/10.1038/nmeth.4035>
- Clark FE, Conte MA, Ferreira-Bravo IA, et al (2017) Dynamic sequence evolution of a sex-Associated b chromosome in lake Malawi cichlid fish. *J Hered* 108:53–62. <https://doi.org/10.1093/jhered/esw059>
- Clark FE, Conte MA, Kocher TD (2018) Genomic characterization of a B chromosome in lake Malawi cichlid fishes. *Genes (Basel)* 9:. <https://doi.org/10.3390/genes9120610>
- Clark FE, Kocher TD (2019) Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish
- Coan RLB, Martins C (2018) Landscape of transposable elements focusing on the B chromosome of the cichlid fish *astatotilapia latifasciata*. *Genes (Basel)* 9:. <https://doi.org/10.3390/genes9060269>
- Conesa A, Götz S, García-Gómez JM, et al (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Conte MA, Gammerdinger WJ, Bartie KL, et al (2017) A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* 18:341. <https://doi.org/10.1186/s12864-017-3723-5>
- Conte MA, Kocher TD (2015) An improved genome reference for the African cichlid, *Metriaclicma zebra*. *BMC Genomics* 16:. <https://doi.org/10.1186/s12864-015-1930-5>
- Cremer T, Cremer C (2006) Rise, fall and resurrection of chromosome territories: A historical perspective Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem* 50:223–272. <https://doi.org/10.4081/995>.
- Cremer T, Cremer M (2010) Chromosome territories. *Cold Spring Harb Perspect Biol* 2:. <https://doi.org/10.1101/cshperspect.a003889>
- Croft JA, Bridger JM, Boyle S, et al (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 145:1119–1131. <https://doi.org/10.1083/jcb.145.6.1119>
- De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 22:1269–1271. <https://doi.org/10.1093/bioinformatics/btl097>
- De Silva DMZA, Utsunomia R, Ruiz-Ruano FJ, et al (2017) High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci Rep* 7:12726. <https://doi.org/10.1038/s41598-017-12939-7>
- Deakin JE, Potter S, O'Neill R, et al (2019) Chromosomics: Bridging the Gap between Genomes and Chromosomes. *Genes (Basel)* 10:. <https://doi.org/10.3390/genes10080627>
- Dixon JR, Selvaraj S, Yue F, et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380. <https://doi.org/10.1038/nature11082>

- Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Douglas RN, Birchler JA (2017) *B chromosomes*. Academic Press, London
- Duan Z, Andronescu M, Schutz K, et al (2012) A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. *Methods* 58:277–288. <https://doi.org/10.1016/j.ymeth.2012.06.018>
- Dudchenko O, Batra SS, Omer AD, et al (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* (80-) 356:92–95. <https://doi.org/10.1126/science.aal3327>
- Duílio DMZ, Pansonato-Alves JC, Utsunomia R, et al (2014) Delimiting the origin of a B chromosome by FISH mapping, chromosome painting and DNA sequence analysis in *astyanax paranae* (teleostei, characiformes). *PLoS One* 9:. <https://doi.org/10.1371/journal.pone.0094896>
- Durand NC, Robinson JT, Shamim MS, et al (2016) Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* 3:99–101. <https://doi.org/10.1016/j.cels.2015.07.012>
- Egecioglu D, Brickner JH (2011) Gene positioning and expression. *Curr Opin Cell Biol* 23:338–345. <https://doi.org/10.1016/j.ceb.2011.01.001>
- Eid J, Fehr A, Gray J, et al (2009) Real-time DNA sequencing from single polymerase molecules. *Science* (80-) 323:133–138. <https://doi.org/10.1126/science.1162986>
- Eklom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl* 7:1026–1042. <https://doi.org/10.1111/eva.12178>
- Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol* 29:51–63. <https://doi.org/10.1016/j.tree.2013.09.008>
- Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol* 29:51–63. <https://doi.org/10.1016/j.tree.2013.09.008>
- Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:. <https://doi.org/10.1186/1471-2105-9-18>
- Emms DM, Kelly S (2019) OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238. <https://doi.org/10.1186/s13059-019-1832-y>
- Ercan S (2014) Mechanisms of X Chromosome Dosage Compensation. *J Genomics* 3:1–19. <https://doi.org/10.7150/jgen.10404>
- Etchegaray E, Naville M, Volff JN, Haftek-Terreau Z (2021) Transposable element-derived sequences in vertebrate development. *Mob DNA* 12:. <https://doi.org/10.1186/s13100-020-00229-5>
- Fan S, Elmer KR, Meyer A (2012) Genomics of adaptation and speciation in cichlid fishes: Recent advances and analyses in African and neotropical lineages. *Philos Trans R Soc B Biol Sci* 367:385–394. <https://doi.org/10.1098/rstb.2011.0247>
- Fantinatti BEA, 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 Genet* 17:10 1186 12863–016–0427–9. <https://doi.org/10.1186/s12863-016-0427-9>
- Fantinatti BEA, Mazzuchelli J, Valente GT, et al (2011) Genomic content and new insights on the origin of the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genetica* 139:1273–1282. <https://doi.org/10.1007/s10709-012-9629-x>
- Feldberg E, Bertollo LAC (1984) Discordance in chromosome number among somatic and gonadal tissue cells of *Gymnogeophagus balzanii* (Pisces: Cichlidae). *Rev Bras Genet* 4:639–645
- Feldberg E, Porto JIR, Alves-Brinn MN, et al (2004) B chromosomes in Amazonian cichlid species. *Cytogenet Genome Res* 106:195–198. <https://doi.org/10.1159/000079287>

- Flynn JM, Hubley R, Goubert C, et al (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* 117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>
- Fraïsse C, Picard MAL, Vicoso B (2017) The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nat Commun* 8:. <https://doi.org/10.1038/s41467-017-01663-5>
- Fudenberg G, Pollard KS (2019) Chromatin features constrain structural variation across evolutionary timescales. *Proc Natl Acad Sci U S A* 116:2175–2180. <https://doi.org/10.1073/pnas.1808631116>
- Furman BLS, Metzger DCH, Darolti I, et al (2020) Sex Chromosome Evolution: So Many Exceptions to the Rules. *Genome Biol Evol* 12:750–763. <https://doi.org/10.1093/gbe/evaa081>
- Gammerdinger WJ, Conte MA, Acquah EA, et al (2014) Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics* 15:. <https://doi.org/10.1186/1471-2164-15-975>
- Gammerdinger WJ, Conte MA, Baroiller JF, et al (2016) Comparative analysis of a sex chromosome from the blackchin tilapia, *Sarotherodon melanotheron*. *BMC Genomics* 17:. <https://doi.org/10.1186/s12864-016-3163-7>
- Gammerdinger WJ, Conte MA, Sandkam BA, et al (2018) Novel Sex Chromosomes in 3 Cichlid Fishes from Lake Tanganyika. *J Hered* 109:489–500. <https://doi.org/10.1093/jhered/esy003>
- Gammerdinger WJ, Kocher TD (2018) Unusual diversity of sex chromosomes in african cichlid fishes. *Genes (Basel)* 9:1. – 20. <https://doi.org/10.3390/genes9100480>
- Ge SX, Jung D, Jung D, Yao R (2020) ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36:2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Gel B, Díez-Villanueva A, Serra E, et al (2016) RegioneR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32:289–291. <https://doi.org/10.1093/bioinformatics/btv562>
- Ghavi-Helm Y, Jankowski A, Meiers S, et al (2019) Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* 51:1272–1282. <https://doi.org/10.1038/s41588-019-0462-3>
- Ghurye J, Rhie A, Walenz BP, et al (2019) Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol* 15:. <https://doi.org/10.1371/journal.pcbi.1007273>
- Gibcus JH, Dekker J (2013) The Hierarchy of the 3D Genome. *Mol Cell* 49:773–782. <https://doi.org/10.1016/j.molcel.2013.02.011>
- Gibcus JH, Samejima K, Goloborodko A, et al (2018) A pathway for mitotic chromosome formation. *Science (80-)* 359:. <https://doi.org/10.1126/science.aao6135>
- Glasauer SMK, Neuhauss SCF (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics* 289:1045–1060. <https://doi.org/10.1007/s00438-014-0889-2>
- Goel M, Sun H, Jiao WB, Schneeberger K (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* 20:277. <https://doi.org/10.1186/s13059-019-1911-0>
- Graphodatsky AS, Kukekova A V., Yudkin D V., et al (2005) The proto-oncogene C-KIT maps to canid B-chromosomes. *Chromosom Res* 13:113–122. <https://doi.org/10.1007/s10577-005-7474-9>
- Green M, Sambrook J *Molecular Cloning: A Laboratory Manual*, 4th edn. Cold Spring Harbor Laboratory Press, New York
- Grewal SIS, Jia S (2007) Heterochromatin revisited. *Nat Rev Genet* 8:35–46. <https://doi.org/10.1038/nrg2008>
- Guiglielmoni N, Houtain A, Derzelle A, et al (2021) Overcoming uncollapsed haplotypes in long-read assemblies of non-model organisms. *BMC Bioinformatics* 22:. <https://doi.org/10.1186/s12859-021-04118-3>
- Guo C, Ito SI, Yoneda M, et al (2021) Fish Specialize Their Metabolic Performance to Maximize Bioenergetic Efficiency in Their Local Environment: Conspecific Comparison Between Two Stocks of Pacific Chub Mackerel (*Scomber japonicus*). *Front Mar Sci* 8:. <https://doi.org/10.3389/fmars.2021.613965>

- Han Y, Wessler SR (2010) MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res* 38:199. <https://doi.org/10.1093/nar/gkq862>
- Hara Y, Tatsumi K, Yoshida M, et al (2015) Optimizing and benchmarking de novo transcriptome sequencing: From library preparation to assembly evaluation. *BMC Genomics* 16:977. <https://doi.org/10.1186/s12864-015-2007-1>
- Hoff KJ, Lange S, Lomsadze A, et al (2016) BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767–769. <https://doi.org/10.1093/bioinformatics/btv661>
- Hong ZJ, Xiao JX, Peng SF, et al (2020) Novel B-chromosome-specific transcriptionally active sequences are present throughout the maize B chromosome. *Mol Genet Genomics* 295:313–325. <https://doi.org/10.1007/s00438-019-01623-2>
- Hong ZJ, Xiao JX, Peng SF, et al (2020) Novel B-chromosome-specific transcriptionally active sequences are present throughout the maize B chromosome. *Mol Genet Genomics* 295:313–325. <https://doi.org/10.1007/s00438-019-01623-2>
- Houben A (2017) B chromosomes – a matter of chromosome drive. *Front Plant Sci* 8. <https://doi.org/10.3389/fpls.2017.00210>
- Houben A, Banaei-Moghaddam AM, Klemme S, Timmis JN (2014) Evolution and biology of supernumerary B chromosomes. *Cell Mol Life Sci* 71:467–478. <https://doi.org/10.1007/s00018-013-1437-7>
- Houben A, Jones N, Martins C, Trifonov V (2019) Evolution, composition and regulation of supernumerary B chromosomes. *Genes (Basel)* 10:161. <https://doi.org/10.3390/genes10020161>
- Huang W, Du Y, Zhao X, Jin W (2016) B chromosome contains active genes and impacts the transcription of A chromosomes in maize (*Zea mays* L.). *BMC Plant Biol* 16. <https://doi.org/10.1186/s12870-016-0775-7>
- Huang YH, Peng SF, Lin YP, Cheng YM (2020) The maize B chromosome is capable of expressing microRNAs and altering the expression of microRNAs derived from A chromosomes. *Chromosom Res* 28:129–138. <https://doi.org/10.1007/s10577-019-09620-2>
- Ibrahim DM, Mundlos S (2020) The role of 3D chromatin domains in gene regulation: a multi-faceted view on genome organization. *Curr Opin Genet Dev* 61:1–8. <https://doi.org/10.1016/j.gde.2020.02.015>
- Imakaev M, Fudenberg G, McCord RP, et al (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* 9:999–1003. <https://doi.org/10.1038/nmeth.2148>
- Jain M, Koren S, Miga KH, et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 36:338–345. <https://doi.org/10.1038/nbt.4060>
- Jehangir M, Ahmad SF, Cardoso AL, et al (2019) De novo genome assembly of the cichlid fish *Astatotilapia latifasciata* reveals a higher level of genomic polymorphism and genes related to B chromosomes. *Chromosoma* 128:81–96. doi: 10.1007/s00412-019-00707-7
- Jones N (2017) New species with B chromosomes discovered since 1980. *Nucl* 60:263–281. <https://doi.org/10.1007/s13237-017-0215-6>
- Karamysheva T, Romanenko S, Makunin A, et al (2021) New data on organization and spatial localization of b-chromosomes in cell nuclei of the yellow-necked mouse *apodemus flavicollis*. *Cells* 10. <https://doi.org/10.3390/cells10071819>
- Kasahara M, Naruse K, Sasaki S, et al (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719. <https://doi.org/10.1038/nature05846>
- Katoh K, Misawa K, Kuma KI, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kichigin IG, Lisachov AP, Giovannotti M, et al (2019) First report on B chromosome content in a reptilian species: the case of *Anolis carolinensis*. *Mol Genet Genomics* 294:13–21. <https://doi.org/10.1007/s00438-018-1483-9>
- Kocher TD (2004) Adaptive evolution and explosive speciation: The cichlid fish model. *Nat Rev Genet* 5:288–298. <https://doi.org/10.1038/nrg1316>

- Kocher TD (2004) Adaptive evolution and explosive speciation: The cichlid fish model. *Nat Rev Genet* 5:288–298. <https://doi.org/10.1038/nrg1316>
- Kocher TD, Behrens KA, Conte MA, et al (2022) New Sex Chromosomes in Lake Victoria Cichlid Fishes (*Cichlidae: Haplochromini*). *Genes (Basel)* 13:804. <https://doi.org/10.3390/genes13050804>
- Kociucka B, Sosnowski J, Kubiak A, et al (2013) Three-dimensional positioning of B chromosomes in fibroblast nuclei of the red fox and the Chinese raccoon dog. *Cytogenet Genome Res* 139:243–249. <https://doi.org/10.1159/000348434>
- Komissarov A, Vij S, Yurchenko A, et al (2018) B chromosomes of the asian seabass (*Lates calcarifer*) contribute to genome variations at the level of individuals and populations. *Genes (Basel)* 9:1. – 15. <https://doi.org/10.3390/genes9100464>
- Konings A (1998) Tanganyika cichlids in their native habitat. Cichlid Press, El Paso
- Kratochwil CF, Kautt AF, Nater A, et al (2022) An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes. *Nat Commun* 13:296. <https://doi.org/10.1038/s41467-021-27685-8>
- Krefting J, Andrade-Navarro MA, Ibn-Salem J (2018) Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* 16:. <https://doi.org/10.1186/s12915-018-0556-x>
- Kuderna LFK, Lizano E, Julià E, et al (2019) Selective single molecule sequencing and assembly of a human Y chromosome of African origin. *Nat Commun* 10:. <https://doi.org/10.1038/s41467-018-07885-5>
- Kuroiwa A, Terai Y, Kobayashi N, et al (2014) Construction of chromosome markers from the lake victoria cichlid paralabidochromis chilotes and their application to comparative mapping. *Cytogenet Genome Res* 142:112–120. <https://doi.org/10.1159/000356128>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <https://doi.org/10.1101/gr.1224503>
- Li Y, Jing XA, Aldrich JC, et al (2017) Unique sequence organization and small RNA expression of a “selfish” B chromosome. *Chromosoma* 126:753–768. <https://doi.org/10.1007/s00412-017-0641-x>
- Lieberman-Aiden E, Van Berkum NL, Williams L, et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-)* 326:289–293. <https://doi.org/10.1126/science.1181369>
- Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* 42:. <https://doi.org/10.1093/nar/gku557>
- Lomvardas S, Maniatis T (2016) Histone and DNA Modifications as Regulators of Neuronal Development and Function. *Cold Spring Harb Perspect Biol* 8:a024208. doi: 10.1101/cshperspect.a024208. PMID: 27371659
- Long HK, Prescott SL, Wysocka J (2016) Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167:1170–1187. <https://doi.org/10.1016/j.cell.2016.09.018>
- López-León MD, Neves N, Schwarzbacher T, et al (1994) Possible origin of a B chromosome deduced from its DNA composition using double FISH technique. *Chromosom Res* 2:87–92. <https://doi.org/10.1007/BF01553487>
- Lucchesi JC (1998) Dosage compensation in flies and worms: The ups and downs of X-chromosome regulation. *Curr Opin Genet Dev* 8:179–184. [https://doi.org/10.1016/S0959-437X\(98\)80139-1](https://doi.org/10.1016/S0959-437X(98)80139-1)
- Lupiáñez DG, Kraft K, Heinrich V, et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161:1012–1025. <https://doi.org/10.1016/j.cell.2015.04.004>

- Ma W, Gabriel TS, Martis MM, et al (2017) Rye B chromosomes encode a functional Argonaute-like protein with in vitro slicer activities similar to its A chromosome paralog. *New Phytol* 213:916–928. <https://doi.org/10.1111/nph.14110>
- Machado HE, Jui G, Joyce DA, et al (2014) Gene duplication in an African cichlid adaptive radiation. *BMC Genomics* 15:161. <https://doi.org/10.1186/1471-2164-15-161>
- Maeso I, Acemel RD, Gómez-Skarmeta JL (2017) Cis-regulatory landscapes in development and evolution. *Curr Opin Genet Dev* 43:17–22. <https://doi.org/10.1016/j.gde.2016.10.004>
- Makunin AI, Dementyeva P V., Graphodatsky AS, et al (2014) Genes on B chromosomes of vertebrates. *Mol Cytogenet* 7:. <https://doi.org/10.1186/s13039-014-0099-y>
- Makunin AI, Kichigin IG, Larkin DM, et al (2016) Contrasting origin of B chromosomes in two cervids (Siberian roe deer and grey brocket deer) unravelled by chromosome-specific DNA sequencing. *BMC Genomics* 17:. <https://doi.org/10.1186/s12864-016-2933-6>
- Makunin AI, Romanenko SA, Beklemisheva VR, et al (2018) Sequencing of supernumerary chromosomes of red fox and raccoon dog confirms a non-random gene acquisition by B chromosomes. *Genes (Basel)* 9:405. <https://doi.org/10.3390/genes9080405>
- Malecová B, Morris K V. (2010) Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs. *Curr Opin Mol Ther* 12:214–222
- Malinsky M, Svoldal H, Tyers AM, et al (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol* 2:1940–1955. <https://doi.org/10.1038/s41559-018-0717-x>
- Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Martins C, Cabral-de-Mello DC, Valente GT, et al (2011) Animal genomes under the focus of cytogenetics. In: *Animal Genomes Under the Focus of Cytogenetics*. Nova Science Publisher, Hauppauge, pp 1–160
- Martins C, Jehangir M (2021) A genomic glimpse of B chromosomes in cichlids. *Genes and Genomics* 43:199–208. <https://doi.org/10.1007/s13258-021-01049-4>
- Martins IC, Portella-Castro ALB, Julio HF (1995) Chromosome Analysis of 5 Species of the Cichlidae Family (*Pisces-Perciformes*) from the Parana River. *Cytologia (Tokyo)* 60:223–231. <https://doi.org/10.1508/cytologia.60.223>
- Masui O, Bonnet I, Le Baccon P, et al (2011) Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell* 145:447–458. <https://doi.org/10.1016/j.cell.2011.03.032>
- Mazzuchelli J, Kocher TD, Yang F, Martins C (2012) Integrating cytogenetics and genomics in comparative evolutionary studies of cichlid fish. *BMC Genomics* 13:. <https://doi.org/10.1186/1471-2164-13-463>
- McArthur E, Capra JA (2021) Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet* 108:269–283. <https://doi.org/10.1016/j.ajhg.2021.01.001>
- Mikheenko A, Valin G, Prjibelski A, et al (2016) Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics* 32:3321–3323. <https://doi.org/10.1093/bioinformatics/btw379>
- Minh BQ, Lanfear R, Trifinopoulos NL-TJ, et al Schmidt March 25, 2022 Tutorials and Manual Phylogenomic software by maximum likelihood
- Misteli T, Soutoglou E (2009) The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat Rev Mol Cell Biol* 10:243–254. <https://doi.org/10.1038/nrm2651>
- Nakajima, R. T. (2019). Análises em larga escala de proteínas e construção de redes biológicas com foco em estudos de cromossomos B. 1–20

- Nascimento-Oliveira JI, Fantinatti BEA, Wolf IR, et al (2021) Differential expression of miRNAs in the presence of B chromosome in the cichlid fish *Astatotilapia latifasciata*. BMC Genomics 22:. <https://doi.org/10.1186/s12864-021-07651-w>
- Navarro-Domínguez B, Martín-Peciña M, Ruiz-Ruano FJ, et al (2019) Gene expression changes elicited by a parasitic B chromosome in the grasshopper *Eyprepocnemis plorans* are consistent with its phenotypic effects. Chromosoma 128:53–67. <https://doi.org/10.1007/s00412-018-00689-y>
- Navarro-Domínguez B, Ruiz-Ruano FJ, Cabrero J, et al (2017) Protein-coding genes in B chromosomes of the grasshopper *Eyprepocnemis plorans*. Sci Rep 7:45200 . <https://doi.org/10.1038/srep45200>
- Nie W, Wang J, Perelman P, et al (2003) Comparative chromosome painting defines the karyotypic relationships among the domestic dog, Chinese raccoon dog and Japanese raccoon dog. Chromosom Res 11:735–740. <https://doi.org/10.1023/B:CHRO.0000005760.03266.29>
- Nishimura O, Hara Y, Kuraku S (2017) GVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics 33:3635–3637. <https://doi.org/10.1093/bioinformatics/btx445>
- Nora EP, Lajoie BR, Schulz EG, et al (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485:381–385. <https://doi.org/10.1038/nature11049>
- O'Connor RE, Romanov MN, Kiazim LG, et al (2018) Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. Nat Commun 9:1. <https://doi.org/10.1038/s41467-018-04267-9>
- Oliveira NL, Cabral-De-Mello DC, Rocha MF, et al (2011) Chromosomal mapping of rDNAs and H3 histone sequences in the grasshopper *rhaphmatocerus brasiliensis* (*acrididae*, *gomphocerinae*): Extensive chromosomal dispersion and co-localization of 5S rDNA/H3 histone clusters in the A complement and B chromosome. Mol Cytogenet 4:24–30. <https://doi.org/10.1186/1755-8166-4-24>
- Östergren G (1945) Parasitic nature of extra fragment chromosomes. Bot. Notiser. 2, 157-163. Ovchinnikov,. Bot Not 2:157–163
- Ou S, Chen J, Jiang N (2018) Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res 46:e126. <https://doi.org/10.1093/nar/gky730>
- Ou S, Jiang N (2018) LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol 176:1410–1422. <https://doi.org/10.1104/pp.17.01310>
- Ou S, Jiang N (2019) LTR_FINDER_parallel: Parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. Mob DNA 10:48. <https://doi.org/10.1186/s13100-019-0193-0>
- Ou S, Su W, Liao Y, et al (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol 20:275. <https://doi.org/10.1186/s13059-019-1905-y>
- Pansonato-Alves JC, Serrano ÉA, Utsunomia R, et al (2014) Single origin of sex chromosomes and multiple origins of B chromosomes in fish genus *Characidium*. In: PLoS ONE. P 9. <https://doi.org/10.1371/journal.pone.0107169>
- Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. <https://doi.org/10.1093/bioinformatics/btm071>
- Paulsen J, Liyakat Ali TM, Collas P (2018) Computational 3D genome modeling using Chrom3D. Nat Protoc 13:1137–1152. <https://doi.org/10.1038/nprot.2018.009>
- Penso-Dolfín L, Man A, Mehta T, et al (2020) Analysis of structural variants in four African cichlids highlights an association with developmental and immune related genes. BMC Evol Biol 20:. <https://doi.org/10.1186/s12862-020-01629-0>
- Perazzo GX, Noleto RB, Vicari MR, et al (2018) B chromosome polymorphism in South American cichlid. Neotrop Biodivers 4:3–9. <https://doi.org/10.1080/23766808.2018.1429164>

- Pires LB, Sampaio TR, Dias AL (2015) Mitotic and meiotic behavior of B chromosomes in *crenicichla lepidota*: New report in the family cichlidae. *J Hered* 106:289–295. <https://doi.org/10.1093/jhered/esv007>
- Poletto AB, Ferreira IA, Cabral-de-Mello DC, et al (2010) Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet* 11:. <https://doi.org/10.1186/1471-2156-11-50>
- Poletto AB, Ferreira IA, Martins C (2010) The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 18S rRNA gene copies. *BMC Genet* 11:. <https://doi.org/10.1186/1471-2156-11-1>
- Poonperm R, Takata H, Hamano T, et al (2015) Chromosome Scaffold is a Double-Stranded Assembly of Scaffold Proteins. *Sci Rep* 5:. <https://doi.org/10.1038/srep11916>
- Pratt HT (1994) Some milestones in the history of bleaching. *Text Chem Color* 26:21–27
- Pryszcz LP, Gabaldón T (2016) Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* 44:e113. <https://doi.org/10.1093/nar/gkw294>
- Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rajičić M, Romanenko SA, Karamysheva T V., et al (2017) The origin of B chromosomes in yellownecked mice (*Apodemus flavicollis*)-Break rules but keep playing the game. *PLoS One* 12:3. <https://doi.org/10.1371/journal.pone.0172704>
- Rambaut A, Drummond AJ, Xie D, et al (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904. <https://doi.org/10.1093/sysbio/syy032>
- Ramírez F, Bhardwaj V, Arrigoni L, et al (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* 9:. <https://doi.org/10.1038/s41467-017-02525-w>
- Ramos É, Cardoso AL, Brown J, et al (2017) The repetitive DNA element BncDNA, enriched in the B chromosome of the cichlid fish *Astatotilapia latifasciata*, transcribes a potentially noncoding RNA. *Chromosoma* 126:313–323. <https://doi.org/10.1007/s00412-016-0601-x>
- Rao SSP, Huntley MH, Durand NC, et al (2015) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 162:687–688. <https://doi.org/10.1016/j.cell.2015.07.024>
- Rhie A, McCarthy SA, Fedrigo O, et al (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592:737–746. <https://doi.org/10.1038/s41586-021-03451-0>
- Roberts NB, Juntti SA, Coyle KP, et al (2016) Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 17:1. <https://doi.org/10.1186/s12864-016-3177-1>
- Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roman H (1947) Mitotic Nondisjunction in the Case of Interchanges Involving the B-Type Chromosome in Maize. *Genetics* 32:391–409. <https://doi.org/10.1093/genetics/32.4.391>
- Roman H Directed fertilisation in maize. In: *Proceedings of the National Acadamey of Sciences of the USA*. pp 450–454
- Ruan J, Li H (2020) Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 17:155–158. <https://doi.org/10.1038/s41592-019-0669-3>
- Ruban A, Schmutzer T, Wu DD, et al (2020) Supernumerary B chromosomes of *Aegilops speltoides* undergo precise elimination in roots early in embryo development. *Nat Commun* 11:2764. <https://doi.org/10.1038/s41467-020-16594-x>
- Ruiz-Herrera A, Farré-Belmonte M (2021) Mechanisms Driving Karyotype Evolution and Genomic Architecture. *Mech Driv Karyotype Evol Genomic Archit*. <https://doi.org/10.3390/books978-3-0365-0157-4>

- Schwarzer J, Misof B, Tautz D, Schlieven UK (2009) The root of the East African cichlid radiations. *BMC Evol Biol* 9:186. <https://doi.org/10.1186/1471-2148-9-186>
- Seehausen O (2006) African cichlid fish: A model system in adaptive radiation research. *Proc R Soc B Biol Sci* 273:1987–1998. <https://doi.org/10.1098/rspb.2006.3539>
- Servant N, Varoquaux N, Lajoie BR, et al (2015) HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 16:259. <https://doi.org/10.1186/s13059-015-0831-x>
- Shao Y Creating a functional single-chromosome yeast. *Nature* 560:331–335
- Sharbel TF, Green DM, Houben A (1998) B-chromosome origin in the endemic New Zealand frog *Leiopelma hochstetteri* through sex chromosome devolution. *Genome* 41:14–22. <https://doi.org/10.1139/g97-091>
- Siefert JL (2009) Defining the mobilome. *Methods Mol Biol* 532:13–27. https://doi.org/10.1007/978-1-60327-853-9_2
- Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Simpson JT, Pop M (2015) The Theory and Practice of Genome Sequence Assembly. *Annu Rev Genomics Hum Genet* 16:153–172. <https://doi.org/10.1146/annurev-genom-090314-050032>
- Smit AF, Hubley R RepeatModeler Open-1.0 (2008); available at www.repeatmasker.org.
- Spielmann M, Lupiáñez DG, Mundlos S (2018) Structural variation in the 3D genome. *Nat Rev Genet* 19:453–467. <https://doi.org/10.1038/s41576-018-0007-0>
- Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res* 32:. <https://doi.org/10.1093/nar/gkh379>
- Stankowski S, Ravinet M (2021) Defining the speciation continuum. *Evolution* (N. Y). 75:1256–1273. <https://doi.org/10.1111/evo.14215>
- Stoebe DM, Dean AM, Dykhuizen DE (2008) The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. *Genetics* 178:1653–1660. <https://doi.org/10.1534/genetics.107.085399>
- Sturmbauer C, Baric S, Salzburger W, et al (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African Lakes. *Mol Biol Evol* 18:144–154. <https://doi.org/10.1093/oxfordjournals.molbev.a003788>
- Su W, Gu X, Peterson T (2019) TIR-Learner, a New Ensemble Method for TIR Transposable Element Annotation, Provides Evidence for Abundant New Transposable Elements in the Maize Genome. *Mol Plant* 12:447–460. <https://doi.org/10.1016/j.molp.2019.02.008>
- Tao H, Li H, Xu K, et al (2021) Computational methods for the prediction of chromatin interaction and organization using sequence and epigenomic profiles. *Brief Bioinform* 22:. <https://doi.org/10.1093/bib/bbaa405>
- Tao W, Xu L, Zhao L, et al (2021) High-quality chromosome-level genomes of two tilapia species reveal their evolution of repeat sequences and sex chromosomes. *Mol Ecol Resour* 21:543–560. <https://doi.org/10.1111/1755-0998.13273>
- Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinforma* 4 10: <https://doi.org/10.1002/0471250953.bi0410s25>
- Terai Y, Morikawa N, Kawakami K, Okada N (2003) The complexity of alternative splicing of hageromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *Proc. Natl. Acad. Sci. U. S. A.* 100:12798–12803
- Valencia JD, Girgis HZ (2019) LtrDetector: A tool-suite for detecting long terminal repeat retrotransposons de-novo. *BMC Genomics* 20:450. <https://doi.org/10.1186/s12864-019-5796-9>
- Valente GT, Conte MA, Fantinatti BEA, et al (2014) Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Mol Biol Evol* 31:2061–2072. <https://doi.org/10.1093/molbev/msu148>

- van Berkum NL, Lieberman-Aiden E, Williams L, et al (2010) Hi-C: A method to study the three-dimensional architecture of genomes. *J Vis Exp*. <https://doi.org/10.3791/1869>
- Van Bocxlaer B, Schultheiß R, Plisnier PD, Albrecht C (2012) Does the decline of gastropods in deep water herald ecosystem change in Lakes Malawi and Tanganyika? *Freshw Biol* 57:1733–1744. <https://doi.org/10.1111/j.1365-2427.2012.02828.x>
- Vietri Rudan M, Hadjur S (2015) Genetic Tailors: CTCF and Cohesin Shape the Genome During Evolution. *Trends Genet* 31:651–660. <https://doi.org/10.1016/j.tig.2015.09.004>
- Vurtture GW, Sedlazeck FJ, Nattestad M, et al (2017) GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* 33:2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>
- Walker BJ, Abeel T, Shea T, et al (2014) Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:. <https://doi.org/10.1371/journal.pone.0112963>
- Ward EJ (1973) The heterochromatic B chromosome of maize: the segments affecting recombination. *Chromosoma* 43:177–186. <https://doi.org/10.1007/BF00483377>
- Wilson EB (1907) The supernumerary chromosomes of Hemiptera. *Science* (80-) 26: 26:26: 870–871.-Open Access Library
- Winkler H (2011) Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche / von Dr. Hans Winkler. Verlag Fischer, Jena
- Wu TD, Watanabe CK (2005) GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859–1875. <https://doi.org/10.1093/bioinformatics/bti310>
- Xiong P, Hulsey CD, Fruciano C, et al (2021) The comparative genomic landscape of adaptive radiation in crater lake cichlid fishes. *Mol Ecol* 30:955–972. <https://doi.org/10.1111/mec.15774>
- Xiong W, He L, Lai J, et al (2014) HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc Natl Acad Sci U S A* 111:10263–10268. <https://doi.org/10.1073/pnas.1410068111>
- Xu Z, Wang H (2007) LTR-FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35:265–268. <https://doi.org/10.1093/nar/gkm286>
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yoshida K, Terai Y, Mizoiri S, et al (2011) B chromosomes have a functional effect on female sex determination in lake victoria cichlid fishes. In: *PLoS Genetics*. P 8. <https://doi.org/10.1371/journal.pgen.1002203>
- Zdobnov EM, Apweiler R (2001) InterProScan - An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848. <https://doi.org/10.1093/bioinformatics/17.9.847>
- Zhou Q, Zhu H mei, Huang Q fei, et al (2012) Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. *BMC Genomics* 13:. <https://doi.org/10.1186/1471-2164-13-109>

8 SUPPLEMENTARY DATA

The supplementary data including supplementary figures and supplementary tables can be accessed using NNNNNNN link.

8.1 List of supplementary Figures

Figure S1. Genome size estimation generated by Genomescope,

Figure S2. Hi-C interaction heatmap

Figure S3. Comparison of *A. latifasciata* chromosome level genome

Figure S4. Insertion age of LTR elements in millions of years (Mya)

Figure S5. Dot-plot comparison between the chr 2 and chr B

Figure S6. Line plots show the chromosome-wise distribution of contact probability

Figure S7. Chromosomes scale heatmap and bars plots of A/B compartments

Figure S8. Hi-C heatmap based on the chromosome-scale assembly of the *A. citrinellus* (a) and *O. niloticus*

Figure S9. Phylogenetic tree and estimated divergence times of cichlid species

Figure S10. Graphical representation of summarized tree of the enriched gene ontology (GO) terms of CAFE expanded genes on B chromosome.

Figure S11. Syntenic dotplots of *A. citrinellus* (left) and *O. niloticus*

Figure S12. Comparative genomic syntenic analysis of *M. zebra* and *A. aureus* genomes

Figure S13. The distribution of rearrangements

Figure S14. Intra-chromosomal inversion

Figure S15. Graph bar plot of normalized (TMM) expression level

Figure S15. Comparison of DEGs between the gonads and muscles tissues

8.2 List of supplementary Tables

Table S1. B_linked genes mapping

Table S2. HiC-Explorer quality controls

Table S3. cichlids_sex_determining_genes_list

Table S4. GenomeScope results

Table S5. Sample sequencing information

Table S6. Assemblies statistic comparison

Table S7. Gene annotation and repeats contents

Table S8. LAI score

Table S9. Sex_linked genes mapping

Table S10. Genome repeat landscape

Table S11. LTRs evolutionary age

Table S12. Cichlids chromosome level assemblies statistic

Table S13. HiC based B-scaffolds

Table S14. B_linked mapped genes hits on chrs

Table S15. TADS_LOOPS_counts_cichlids

Table S16. orthologous_genes_statistic

Table S17. orthogroup_gene_count

Table S18. A. latifascaita_expanded_genes_enrichment

Table S19. A. latifascaita-contracted_genes_enrichment

Table 20. B_expnded_genes_info

Table S21. SV count per chr

Table S22. FPKM_Pvalue_FDR_B-realted_genes

Table S23. Brain_highly_sig_DEGs

Table S24. Enriched_Bchr_genes