



Peer Community In Evolutionary Biology

Young sex chromosomes discovered in white-eye birds

Kateryna Makova based on peer reviews by **Melissa Wilson**, **Gabriel Marais** and 1 anonymous reviewer

Thibault Leroy, Yoann Anselmetti, Marie-Ka Tilak, S everine B erard, Laura Csukonyi, Ma eva Gabrielli, C eline Scornavacca, Borja Mil a, Christophe Th ebaud, Benoit Nabholz (2019) A bird's white-eye view on neosex chromosome evolution. Missing preprint_server, ver. Missing article_version, peer-reviewed and recommended by Peer Community in Evolutionary Biology. [10.1101/505610](https://doi.org/10.1101/505610)

Submitted: 24 January 2019, Recommended: 11 June 2019

Cite this recommendation as:

Makova, K. (2019) Young sex chromosomes discovered in white-eye birds. *Peer Community in Evolutionary Biology*, 100073. [10.24072/pci.evolbiol.100073](https://doi.org/10.24072/pci.evolbiol.100073)

Published: 11 June 2019

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Recent advances in next-generation sequencing are allowing us to uncover the evolution of sex chromosomes in non-model organisms. This study [1] represents an example of this application to birds of two Sylvioidea species from the genus *Zosterops* (commonly known as white-eyes). The study is exemplary in the amount and types of data generated and in the thoroughness of the analysis applied. Both male and female genomes were sequenced to allow the authors to identify sex-chromosome specific scaffolds. These data were augmented by generating the transcriptome (RNA-seq) data set. The findings after the analysis of these extensive data are intriguing: neoZ and neoW chromosome scaffolds and their breakpoints were identified. Novel sex chromosome formation appears to be accompanied by translocation events. The timing of formation of novel sex chromosomes was identified using molecular dating and appears to be relatively recent. Yet first signatures of distinct evolutionary patterns of sex chromosomes vs. autosomes could be already identified. These include the accumulation of transposable elements and changes in GC content. The changes in GC content could be explained by biased gene conversion and altered recombination landscape of the neo sex chromosomes. The authors also study divergence and diversity of genes located on the neo sex chromosomes. Here their findings appear to be surprising and need further exploration. The neoW chromosome already shows unique patterns of divergence and diversity at protein-coding genes as compared with genes on either neoZ or autosomes. In contrast, the genes on the neoZ chromosome do not display divergence or diversity patterns different from those for autosomes. This last observation is puzzling and I believe should be explored in further studies. Overall, this study significantly advances our knowledge of the early stages of sex chromosome evolution in vertebrates, provides an example of how such a study could be

conducted in other non-model organisms, and provides several avenues for future work.

References:

[1] Leroy T., Anselmetti A., Tilak M.K., Bérard S., Csukonyi L., Gabrielli M., Scornavacca C., Milá B., Thébaud C. and Nabholz B. (2019). A bird's white-eye view on neo-sex chromosome evolution. bioRxiv, 505610, ver. 4 peer-reviewed and recommended by PCI Evolutionary Biology. doi: [10.1101/505610](<https://dx.doi.org/10.1101/505610>)

Reviews

Evaluation round #2

DOI or URL of the preprint: [10.1101/505610](https://doi.org/10.1101/505610)

Version of the preprint: 3

Authors' reply, 12 May 2019

Dear Kateryna Makova,

I have just uploaded the new version of your manuscript on BioRxiv. <https://www.biorxiv.org/content/10.1101/505610v3>

All the requested changes have been made.

Thank you very much for your work and time, as well as those of the three reviewers, who helped us to substantially improve the manuscript.

Best regards, Thibault

Decision by [Kateryna Makova](#), posted 02 May 2019

Please revise the manuscript according to my comments

I have attached the annotated file with my comments. Additionally, please: - make sure you leave space between a value and "kbp" or "Mb"; - italicize all scientific names in the paper; - spell out all numbers under 10; - add statistical test used prior to mentioning every P value [Download recommender's annotations](#)

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/505610>

Version of the preprint: 2

Authors' reply, 26 April 2019

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Kateryna Makova](#), posted 18 March 2019

Please revise the manuscript according to reviewers' comments

Additional requirements of the managing board:

As indicated in the 'How does it work?' section and in the code of conduct, please make sure that:

-Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad (to pay) or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.

-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.

-Details on experimental procedures are available to readers in the text or as appendices.

-Authors have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this preprint declare that they have no financial conflict of interest with the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the PCI Evol Biol recommenders."

Reviewed by [Melissa Wilson](#), 30 January 2019

Overall, I think the topic of the manuscript is very interesting, and is why I agreed to review. For now, however, I am cautious with any interpretations. I am waiting to evaluate the rest of the manuscript until several questions about the methodology are answered. Some are strait-forward, but just useful to have in the same place, and for clarity. Others are essential for interpreting measurements of diversity and the transcriptome analyses.

Methods questions 1. The authors start by describing sampling from one female *Z. borbonicus* and extracting DNA from tissue - which tissue?

1. Which tissue did the authors collect DNA from from the *Zosterops pallidus*? Please specify the sex of the individual in the methods, currently it just says, "individual".
2. The authors refer to liver and/or muscle - did they do multiple extractions for each individual? Given there are only two birds, please be specific about what was collected, how many extractions were done, whether these extractions were pooled, and what exactly (tissue sample and species) was sequenced.
3. Were the samples multiplexed? Or were each run on an individual lane? What sequencing depth was conducted for each of the samples?
4. For the 10X PacBio sequencing coverage, how many SMRT cells were used (or what was the technology) to obtain the 10X coverage? Is this a theoretical 10X coverage? How much was the coverage when aligned to the reference genome, and how did it vary on the Z and autosomes?
5. For *Z. barbonicus*, describing the RNA extraction, the authors say they extracted RNA then stabilized it. I was under the impression RNA would be stabilized in a solution (like RNAlater), then extracted later. If it really was extracted immediately in the field, that is worth noting. Else, can you describe how the tissues were collected for RNA isolation - how long was the individual dead? Did you have RNAlater in the field? Or did you bring the individual to the lab to dissect?
6. Did you get RIN values for each of the samples before sequencing? If so, please report them.
7. What kind of RNA library prep was performed? Did the authors do ribosomal RNA (rRNA) depletion, or was total RNA sequenced? If total RNA, what kind of effect does this have on the ability to detect transcripts and what proportion of transcripts were assembled/aligned to ribosomal sequences (and how many thus were used for annotation)?

8. How many millions of reads were sequenced for each sample? The authors refer to (1 line) of Hiseq2500 sequencing. I think they mean, (1 lane)? If so, how many predicted reads per sample, and how were the samples multiplexed (or were they)?
9. Can the authors comment on if (and then if so) why the two RNA extracts were pooled prior to sequencing rather than sequencing them separately and combining the fastq files afterwards? Each tissue is presumed to have very different genes expressed in them, and even different transcripts.
10. Will the authors please provide the statistics and fastqc reports for the data pre- and post-trimming (or at least post-trimming) for the DNA and RNA sequence samples?
11. For the DNA, the quality score threshold used was 15 (meaning somewhere between 1/10 and 1/100 of the bases in the DNA are likely to be errors). For the RNA, the quality score threshold used was just 5, meaning - as I understand PHRED scores - more than 1/10 base pairs in the RNA are expected to be errors. Perhaps this was because the authors only wanted to use transcripts to identify transcribed regions, but this low of a PHRED quality score is concerning, as it is a probability that nearly a third of base positions are incorrect in the RNA data. I would like to see more assessment of the quality of the RNAseq, and the effects of using such low quality RNA in assembling a transcriptome. Presumably there is a lot of degraded RNA, and this will make both transcriptome assembly, and gene expression analyses quite difficult. But, perhaps the authors have evidence that such a low PHRED threshold for RNAseq data still yields trustworthy transcriptome annotation?
12. The authors use mean depth to identify autosomal, Z, and W-linked scaffolds, but this should not be the only evaluation, as depth in sequencing experiments is not evenly distributed across the genome, and especially in small scaffolds can confound estimates of Z, and W linkage. At least one additional measure (heterozygosity for example), should be applied to further support the Z and W linkage. Also, have the authors considered the effects of Z-W homology? In an effort to be conservative, it seems like they are identifying regions that are highly differentiated between Z and W, and as such, this should be acknowledged.
13. When the authors describe measuring genome-wide genetic diversity, the description of how those BAM files were produced is not included, but should be (albeit briefly) here. If the authors used the same trimming and quality thresholds for the DNA used in these analyses, I would be cautious about interpreting the patterns of diversity. Further, are these BAM files from DNA or RNA, it isn't specified. If they are from RNA, there are additional challenges. If they are from DNA, it is helpful to know what depth, how it varies on Z and autosomes, and how this coverage will affect variant calling.

Reviewed by **Gabriel Marais**, 30 January 2019

Leroy et al.'s ms is focused on a neo-sex chromosome system in birds. They have sequenced 2 bird genomes and identified the scaffolds from the sex and neo-sex chromosomes. They then performed a thorough characterization of this system using comparative genomics, population genetics, molecular evolution and phylogenetics approaches. They found that the neo-sex chromosomes are recent (1-3.5 My old). They originated from a partial autosomal fusion of which they identified precisely the breakpoints using an outgroup bird genome. The neo-sex chromosomes harbour hallmarks of early sex chromosome evolution: reduced diversity (on both neo-Z and neo-W, stronger on the latter), higher dN/dS on the neo-W genes and accumulation of TEs on the neo-W.

In vertebrates, most of the studies on sex chromosomes have focused on old systems such as the mammalian XY and the avian ZW chromosomes. Very little is known on the early stages of sex chromosome evolution. In invertebrates, the study of neo-sex chromosomes (in drosophila, see for example papers from D. Bachtrog and AB Carvalho groups) has provided many key observations to understand the early steps of sex chromosome

evolution and Y degeneration. To my knowledge, this is the first ms doing similar work on a vertebrate neo-sex chromosome system. Although no completely unexpected patterns have come out of this study, it filled an important gap in the literature. We have data on a vertebrate neo-sex chromosome system showing that neo-W ongoing degeneration resembles what has been observed in neo-Y systems in drosophila.

Another positive point about this ms is the amount of data and the number of analyses that have been done. Two genomes have been sequenced, one of which is of very high quality. Many analyses have been done (reconstructing the autosome-sex chromosome fusion, identifying the sex and neo-sex scaffolds, dating the system, computing the genetic diversity, looking at the dN/dS, the TE content, the GC content, etc...). It is a lot of data and findings in a single paper. For most analyses, the results are very neat (e.g figures 2, 4, 5, 8, 9). In other systems, several papers have been published to report a similar work. On the other hand, the ms reads very well and is clear and interesting all along.

I am thus very positive about this ms and I think it would merit a PCI evol biol recommendation. I only have minor comments/suggestions, which are listed below.

1) I think the title could be improved. At present it does not tell a lot about what's in the ms. I think a title such as "The young neo-sex chromosome system in *Zosterops* provides insight on the early steps of ZW chromosome evolution in birds" or something similar (including the key words neo-sex chromosome and early steps of ZW chromosome evolution or ongoing W degeneration) should be preferred.

2) In the literature, the term neo-sex chromosome is restricted to the part that has been fused to the original sex chromosomes (see for example Bachtrog 2013 figure 2). In Leroy et al. they call the neo-sex chromosomes the fused autosomal part + the original sex chromosomes (see figure 4 and throughout the ms). I think this needs to be either corrected or at least acknowledged, to avoid readers getting confused.

3) The authors may want to cite work done on plant neo-sex chromosomes (and speciation) such as Weingartner & Delph 2014.

4) Perhaps, it would be interesting to speculate on why such an A-ZW fusion has taken place. Is there something special about *Zosterops*, in terms for example of sexual dimorphism (one possible driver of neo-sex chromosome evolution, see Kitano et al. 2009).

References

Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet.* 2013 Feb;14(2):113-24.

Weingartner LA, Delph LF. Neo-sex chromosome inheritance across species in *Silene* hybrids. *J Evol Biol.* 2014 Jul;27(7):1491-9.

Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, Kingsley DM, Peichel CL. A role for a neo-sex chromosome in stickleback speciation. *Nature.* 2009 Oct 22;461(7267):1079-83.

Reviewed by anonymous reviewer 1, 30 January 2019

The manuscript "A bird's white-eye view on avian sex chromosome evolution" by Leroy et al. reports a high-quality genome assembly of the "great speciator", the songbird genus *Zosterops*. This group of species may have the fastest rate of speciation after East African cichlids. The authors use the chromosome-level information of this genome assembly and additional *Zosterops* species plus outgroups to convincingly infer a) the timing of *Zosterops* diversification, b) the evolution of remarkable neo-sex chromosome of sylvioid songbirds through fusion of the Z and W chromosomes with half of chromosome 4A, c) the molecular evolution of genes since neo-sex chromosome emergence, and d) the accumulation of transposable elements (TEs) since neo-sex chromosome emergence.

My impression is that this manuscript is a well-written and comprehensive genome analysis of an interesting study system. The methods appear sound and the main conclusions are well supported by the data. I also commend the authors for their attention to detail in the methods section, and for providing all scripts and datasets on figshare. However, I have some suggestions for how to further improve the manuscript. Most of

these are related to clarification of specific statements or methods as a service to the reader.

Main comments:

1. Throughout the text: I wonder why the authors use the term “translocation” instead of “fusion” for the mechanism leading to a neo-sex chromosome? Please explain this choice at the first mentioning of the Sylvioidea neo-sex chromosome system, or replace with the term “fusion”. As far as I am aware of, other papers on this system usually use the term “fusion”.
2. Throughout the text/figures: Please always add the term “contig” or “scaffold” in context of N50. Scaffold N50 and contig N50 are quite different metrics and should therefore be explicitly mentioned.
3. Throughout the text: The authors mention a megabase-scale deletion on the neoW-4A gametolog. How many genes are located in this region in the neoZ-4A gametologous region? It might be interesting to state the number of genes lost in this deletion event, highlighting that a single large intrachromosomal rearrangement can lead to significant differences between Z/W gametologous regions.
4. Throughout the text: Strictly spoken, gametologs are not paralogs (those arose through gene duplication) but homologs that arose through sex chromosome differentiation.
5. Lines 171-190, Figure 1, Figure S2: It is not clear to me what the TE content in genes refers to – are these TEs nested inside introns/exons/UTRs or even part of the protein-coding sequence? Figure S2 shows that some genes have a TE content of nearly 100%, so I wonder if these could be either TE-derived genes or gene models incorporating genes? Please clarify.
6. Lines 226-234, lines 658-667: I commend the authors for improving the assemblies of 25 other bird species with DeCoStar and making these data freely available on figshare. The question is: How informative are these – do the authors have thoughts on whether these assemblies might be biased towards those assemblies with higher quality and potentially underestimating the number of intra/interchromosomal rearrangements? Please briefly comment on these possibilities.
7. Line 409 (Figure 4): a) How are the LASTZ hits ordered on the Y axis? What about regions with hits to other chromosomes – are these translocated regions from autosomes or spurious LASTZ hits due to repeats? b) Is it known where the pseudoautosomal region (PAR) is on the neoZ and neoW chromosomes? If this is the case, please indicate the chromosome ends containing the PAR in this figure.
8. Lines 482-485: Have the authors looked into whether there are “heterozygous” sites in those neoW-W windows with higher nucleotide diversity? Could some ambiguous mapping to repeat regions on neoW-W play a role here?
9. Lines 556-557: Some of the high dN/dS values might be due to positive selection instead of “ongoing pseudogenization”. Did the authors check for premature stop codons and/or frameshifts in these genes?
10. Lines 625-626: Note that contig N50 might be an even better predictor for assembly quality in this context.
11. Lines 831-832: Please explain this statement. Shouldn't a fission of a chromosome into two parts (no matter how close/distant to the centromere) still change the overall recombination rate?
12. Lines 879-882: I fear obtaining a “full sequence” of any W chromosome is not feasible as long as read lengths are below the megabase-scale. I suggest using the term “high-quality sequence” instead.
13. Lines 941-942: Similar TE densities were also reported for a partial W sequence of white-throated sparrow (Davis, J.K., P.J. Thomas & J.W. Thomas. 2010. A W-linked palindrome and gene conversion in New World sparrows and blackbirds. *Chromosome Res.* 18: 543–553.).
14. Lines 948-950: Similar to zebra finch, collared flycatcher also appears to have recently active LTR elements (Suh, A., L. Smeds & H. Ellegren. 2018. Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. *Mol. Ecol.* 27: 99–111).
15. Lines 951-952: Is this a “recent burst of LTR elements on autosomes” or rather an overall burst of LTR elements with an increased retention of insertions in low-recombination regions, such as the neoW-4A region?
16. Lines 1485-1486: Why were Repbase repeats used only from chicken and not also those from zebra finch (i.e., by selecting taxon “Aves”)? As the authors state in the manuscript, there was recent LTR activity in zebra

finch and it is possible that zebra finch might be sharing some TEs with *Zosterops*. This might possibly reduce the percentage of “No Category” (i.e., “unknown” or “unclassified”) repeats in the *Zosterops* annotation.

Additional minor comments:

17. Lines 4-5: I assume this statement applies to the recombining sex chromosome (Z or X), but not the sex-limited one (W and Y)? Or does this statement only apply to synteny of sex chromosomes (but not to intrachromosomal rearrangements)? Please clarify.
18. Line 27: The term “transposed regions” is unclear here. How about using the term “sex-linked”, to keep it independent of the mechanism leading to sex linkage/sex chromosome differentiation (e.g., transposition, translocation, fusion)?
19. Lines 44-45: I assume also purifying selection should be more efficient in the heterogametic sex.
20. Line 52: The authors could use the more general term “linked selection” instead of “background selection” here.
21. Line 65: I suggest replacing the term “all birds” with “most birds” here.
22. Line 118: I suggest replacing the term “chromosomes” with “pseudochromosomes” here.
23. Line 135: Please replace “surprising” with “surprisingly”.
24. Line 244: This approach not only assumes “perfect synteny” but also “collinearity”, which I suggest adding here.
25. Line 319: I suggest adding “onset of the” before “diversification”.
26. Line 422: It is not clear to me how “high activity of transposable elements” can lead to homology (only) between this scaffold and chromosome 4A. Please clarify.
27. Line 579: Please add “p=” before “0.096”.
28. Line 638: Are GC-rich regions mainly difficult because of assembly issues or rather underrepresentation in sequencing data?
29. Line 807-808: I assume the authors mean GC content negatively correlates with chromosome size?
30. Line 813-814: In addition to “instantaneous changes in chromosome sizes”, the new recombination regime (male-specific) might also play a role here, assuming that recombination quickly ceased (or was quickly reduced) between neoW-4A and neoZ-4A.
31. Line 912: What do the authors mean with the term “non-recombinant W chromosome”? Please clarify.
32. Line 1044: Typo in “lane”.
33. Lines 1243-1245: If de-novo prediction was done, I assume the authors mean “RepeatModeler” instead of “RepeatMasker” here?
34. Lines 1286-1289: If there was disagreement between scaffold orders, was preference given to LASTZ or DeCoSTAR? Please clarify.
35. Lines 1309-1311: Were these distances to chromosome ends measured in zebra finch chromosomes or *Z. borbonicus* pseudochromosomes?
36. Line 1338: Please remove “that”.
37. Lines 1425-1430: Please clarify how a single base on the neoW chromosome can have non-zero coverage in males.
38. Lines 1467-1478: In contrast to most parts of the manuscript, here the English names for the study species are used. Consider simplifying this as a service to the reader, for example by only using the scientific name.