Whole Genome Sequencing and Assembly of the House Sparrow, *Passer domesticus*

Vikas Kumar*[1], Gopesh Sharma*[2,7], Sankalp Sharma*[2,3], Dalia Vishnudasan*[4], Gopinathan Maheswaran[1], Kaomud Tyagi[1], Inderjeet Tyagi[1], PB Kavi Kishor,[5] Gyaneshwer Chaubey[6] and Prashanth Suravajhala[2,4]*[#]

1. Zoological Survey of India, Kolkata, 700016
2. Bioclues.org, India
3. NIMS University, Jaipur,
4. Systems Genomics Lab, Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Clappana PO 690525, Kerala, India
5. Department of Genetics, Osmania University, Hyderabad 50007, India
6. Cytogenetics Lab, Department of Zoology, Banaras Hindu University, Varanasi, UP, India
7. Department of Biosciences, Manipal University Jaipur, Jaipur

*Equal contributing authors

[#] Correspondence: prash@bioclues.org and pbkavi@yahoo.com

Abstract

The common house sparrow, *Passer domesticus* is a small bird belonging to Passeridae. We provide here a high-quality whole genome sequencing and assembly for the house sparrow, The final genome assembly was made using a Shovill/SPAdes which consisted of contigs spanning 268193 bases. We employed rigorous statistical thresholds to check the coverage as the Passer genome showed considerable similarity with *Gallus gallus* (chicken) and *Taeniopygia guttata* (Zebra finch) genomes providing a functional annotation. This new annotated genome assembly will be a valuable resource as a reference for comparative and population genomic analyses of passerine, avian, and vertebrate evolution

Keywords: Passerines, house sparrow, next generation sequencing, evolution, adaptation

Significance

Avian evolution has been of great interest in the context of extinction. Annotating the genomes such as Passerines would be a significant interest as we could understand the behavior/foraging traits and further explore their evolutionary landscape. In this work, we provide a full genome sequence of Indian house sparrow, *viz. Passer domesticus* which will

serve as a useful resource in understanding the adaptability, evolution, geography, Allee effects and circadian rhythms.

Introduction

Over the last 12 years, a number of bird reference genomes have been studied providing phylogenetic relationships among the avians (Zhang et al. 2012; Jarvis et al. 2014; Poelstra et al. 2014; Frankl-Vilches et al. 2015; Friis et al. 2018; Louha et al. 2019; Peñalba et al. 2019; Ducrest et al. 2020, Wang et al. 2020), The bird genome project (B10K) has been of great significance in providing major scientific breakthroughs in phylogenetics (Zhang et al. 2015, Feng et al. 2020). With more than 1200 species, comprising 13% of all known world's avian, India is blessed considering avian diversity. Regrettably, India is ranked 3rd in having rare and threatened avian species in the world (Rahmani 2008). House Sparrow was introduced to India via Europe from North Africa and Eurasia by the ancient Romans. Sparrows are found in a variety of habitats, including grasslands, forests, deserts, and agricultural areas and in urban areas, such as parks and gardens. They are omnivorous, typically feeding on insects, spiders, worms, seeds, fruits, and grains, and are mainly seed eaters but feed their young ones with insects and other invertebrates; during the breeding periods they prefer areas with a high abundance of invertebrates (Summers-Smith, 2003). Over the years, there has been a tremendous decline in their population worldwide and therefore understanding causal mechanisms through which urbanization affects their population remains limited because of numerous perplexing factors, what is believed to be due to rapid urbanization along with deforestation, lack of cavity nesting, along with absence of hedges in modern landscaping. Many hypotheses have been proposed explaining the house sparrow population decline, for example increased predation by domestic cats or sparrow hawks (*Accipiter nisus*), cleaner streets providing reduced foraging opportunities, competition for food from other urban species, loss of nesting sites, particularly under the eaves and in the roofs of houses, pollution/air quality, both in terms of immediate toxicity and indirect toxicity through the food supply, increased use of pesticides in parks and gardens, disease transmission (Crick et al., 2002) and finally, the Allee effect (Summers-Smith, 2003). *Passer domesticus,* between 1970-1990 had an extremely large breeding population with over 63 million pairs, and has been declining over the years in the European Union including Turkey and Netherlands. Analyzing house sparrow habitat characteristics on a fine scale in almost 200 census sites with census data from 2003-2017, was one of the most complex field studies in Urban Paris, with a dramatic decline of ~89% of the

species over the study period observed (De Coster et al. 2015; Mohring et al. 2021). In India, a sharp decline in the house sparrow population was observed across Mumbai, Bengaluru, Hyderabad, and other major cities with certain places in India having over 70% decline (Gulati, 2005). Typically the lifespan of the house sparrow is 3 - 5 years in the wild and only about 20 percent of young live past their first year. Cold weather and food availability decide the longevity of Sparrows.*Plasmodium relictum*, a parasite infection, has also affected house sparrow demography across suburban London where sparrows have declined by 71% since 1995 (Dadam et al. 2019).

The genome sequencing efforts have been yielding contentious debate heralding short and long read chemistry over the last few years. While the peacock genome has yielded results (Dhar et al. 2019), we have earlier contemplated asking questions on the Passer genome sequencing (Suravajhala and Acharya, 2014). The avian genome project (https://b10k.genomics.cn/ last accessed on September 5, 2023) representing several orders is aimed to resolve the phylogeny and collate data for testing hypotheses, understanding extinction and speciation of the birds besides demographic events, their roles of drift and selection in the divergence process (Sendell-Price et al. 2020). Recently, Magallanes-Alba et al. (2023) have done rapid genome functional annotation pipeline anchored to the house sparrow genome re-annotation and provided transcripts. Our current genome sequencing of house sparrow is based on a muscle of the bird followed by assembly, annotation employing *in silico* approaches using a myriad of tools that could serve as a valuable resource to understand passerines' evolution, bio/ethno/geography and demography (Illera et al. 2018).

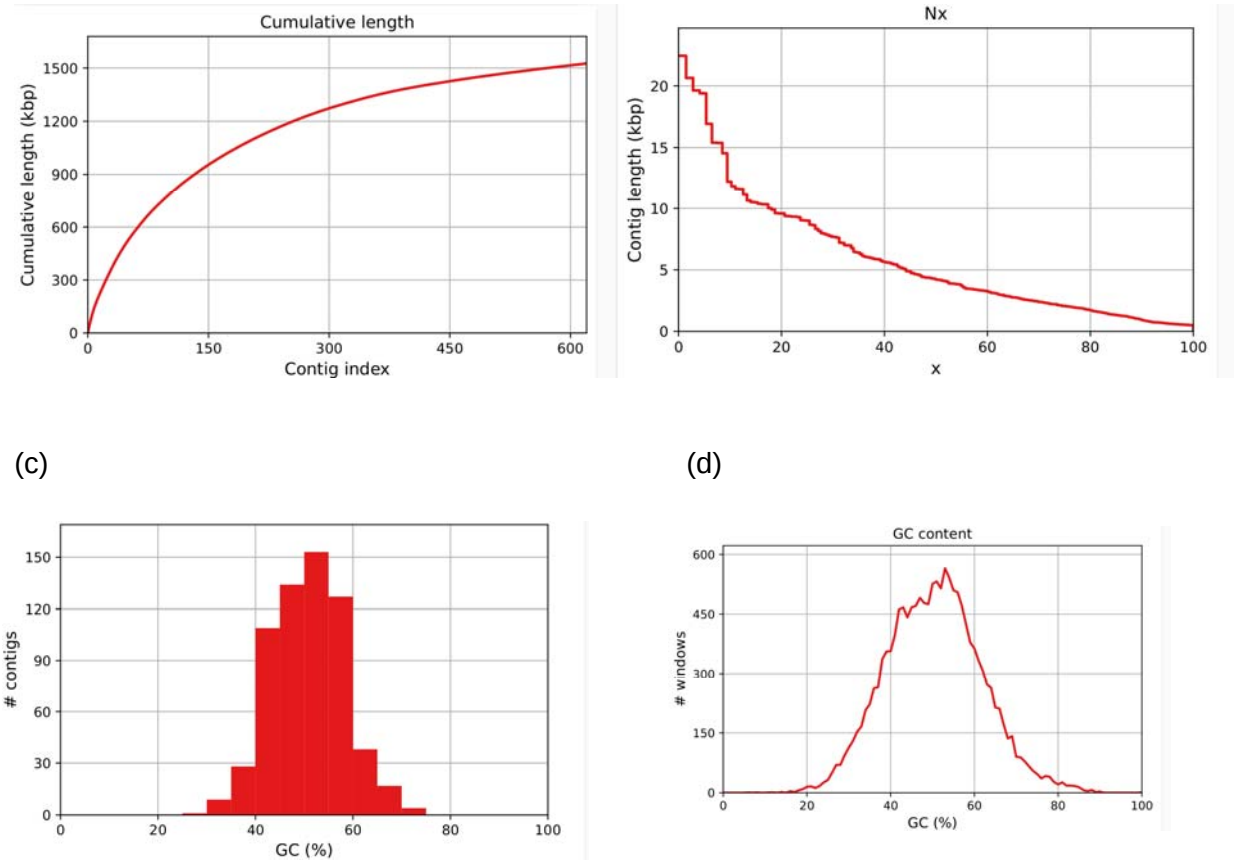
Results

Assembly and Quality Control

The total length obtained by the Shovill/SPAdes was with an estimated k-mer set to 31. The final assembly consisted of contigs of over 1 kb, and a mean N50 of 4233 (Table 1) with a 40826760 x 2 PE reads covering 14 GB each file. We employed rigorous statistical thresholds to check the coverage as the Passer genome showed considerable similarity with *Gallus gallus* (chicken) and *Taeniopygia guttata* (Zebra finch) genomes (Figure 1a and 1b), providing plausible evidence of several chromosome architectures. There are 268193 total bases covered by a set of intervals wherein bases that are covered by more than one interval are not counted twice. In the x-axis, the N50 measures N50 across the assemblies are indicative showing that higher Nx values are more contiguous with the mean GC% of 50.38 (Figure 1c and 1d). We

obtained a total of 710 contigs with the largest contig size of 22436 bp (Table 1). We sought to check the auN which usually falls between N50 and N40, nonetheless, in this case, auN had a significant jump showing 5951 indicating that measure of contiguity is high.

(a) (b)



(c) (d)

Figure 1: (a) Cumulative length vs contig length with (b) N50 (Nx) plotted, (c) GC% against number of contigs and (d) GC content against number of windows

# contigs	620
# contigs (>= 0 bp)	1391
# contigs (>= 1000 bp)	376
# contigs (>= 5000 bp)	76

# contigs (>= 10000 bp)	20
# contigs (>= 25000 bp)	0
# contigs (>= 50000 bp)	0
Largest contig	22436
Total length	1525023
Total length (>= 0 bp)	1696224
Total length (>= 1000 bp)	1365656
Total length (>= 5000 bp)	673668
Total length (>= 10000 bp)	275530
Total length (>= 25000 bp)	0
Total length (>= 50000 bp)	0
N50	4233
N90	925
auN	5951
L50	96
L90	384
GC (%)	50.38
Predicted genes	
# predicted genes (unique)	1282
# predicted genes (>= 0 bp)	1216 + 66 part
# predicted genes (>= 300 bp)	824 + 51 part
# predicted genes (>= 1500 bp)	85 + 5 part
# predicted genes (>= 3000 bp)	17 + 1 part

Table 1: Statistics without reference

An approximate 25% of the genome has sequences matched to bacteria

The Passer genome showed considerable similarity (ca.25%) with bacterial sequences indicating that a large number of gut microbiota may have evented horizontal gene transfer (HGT) over the years (Hotopp, 2011). This could be attributed to the bird's foraging on insects and small grains with bacteria inhabited in them could exist with endosymbionts and their invertebrate hosts, particularly insects. We argue that these could in principle be largely

associated with modern evolutionary theory promoting such gene transfer. In addition, the alignment between these genomes is indicative of inversions/transversions which is in agreement with bird genomes (Stoltzfus and Norris, 2016). The 710 contigs were mapped to the GenBank yielding 152 hypothetical proteins indicating that a large amount of proteins emerged from the known unknown regions, implying the need to annotate the genome significantly. Among these, 252 sequences from protein domains match to animals while the rest match to bacteria. We observed a large number of Kelch domains in addition to a PHD finger 1 found in Histone-lysine N-methyltransferase 2C (KMT2C) and 2D (KMT2D) protein that is largely conserved. Also termed myeloid/lymphoid or mixed-lineage leukemia protein 3 (MLL3), it is associated with circadian factors contributing to genome-scale circadian transcription. These observations could indicate that a large number of genes possibly have undergone divergent evolution or the genes might have been lost and found through speciation. This fragmented genome is also indicative of BUSCO statistics from avian genomes that were reported earlier in Chaffinch genomes (Receurda et al. 2021).

Seven superfamily domains are conserved between *Passer domesticus and Acanthisitta chloris*

The final set of contigs obtained for assembly were mapped against NCBI taxa (Supplementary Table 1) even as it matches to 14947 contigs with 769 of them making 100% query coverage with various hits from taxa (Supplementary Table 2). Amongst the superfamily members, intermediate filament protein, BASP1: Brain acid soluble protein 1, glycine rich LPXTG-anchored collagen-like adhesin and collagen with keratin and some hypothetical protein-FTsK translocase domains. What remains intriguing is the presence of probable chromatin-remodeling complex ATPase chain which is possibly associated with Allee effects that occur when individual fitness suffers at low population size or density (Gascoigne, 2003). We also argue that the genetic architecture and community-wide admixture could provide insights into their evolution (Enbody et al. 2023). On the other hand, keeping in view of the sample characteristics as a measure of genome completeness, we also mined the long read data from Oxford nanopore technologies (not shown). While the aforementioned statistics in the form of N50, N90, BUSCO are optimized for short reads, we were short of full potential in analyzing the genomic reads generated from long reads. Furthermore, the sequences generated need to be polished to overcome the mis-assemblies and have accurate scaffolding for generating more suitable short reads.

Avian genomes are ca. 70% smaller than mammalian genomes and Passerines being small are thought to maintain the contiguity that we have shown as compared to other vertebrates birds. We are limited by the amount of genomic attributes coming from non-coding regions, nevertheless, we argue that this could be because of the large number of hypothetical proteins which demand accurate annotation. While many genomes show syntenic, we are indefatigable to sequence more genomes that could provide us insights on adaptation, evolutionary mechanisms, substitutions, and importantly the circadian rhythms or electromagnetic effects that these birds might be associated with, thus drifting away from the urban life.

Conclusions

We provide here a high-quality whole genome sequencing and assembly for the house sparrow, *Passer domesticus* which could serve us to ask a large number of questions from genomics. Our annotation serves as a useful resource to check adaptation, divergence and speciation. The orthologous annotation and protein mapping, bacterial correlates show that the HGT events may not be ruled out.

Material and Methods

Sample collection and genome assembly

A bird wing (muscular tissue) was extracted from a house sparrow male which was found dead in the lawns of the Zoological Survey of India, Kolkata, and frozen immediately in liquid nitrogen. The sample was handled by Unipath Labs for DNA extraction, sequencing and the paired end raw reads were obtained by us for further processing. We performed genome assembly using Shovill/SPAdes (Seemann, 2017). Initially, we mapped *Gallus gallus* and *Taeniopygia guttata* and the alignment was not successful as upon further alignment using bowtie2, we found 99% mapped reads (Figure 2) which led us to conduct downstream scaffolding. The gene completeness for Passer was assessed through Benchmarking Universal Single-Copy Orthologs (BUSCO; Seppey et al. 2019) by using the orthologous genes in the *Gallus gallus* (chicken) genome. A summary of methods can be found in Figure 3

(a) (b)

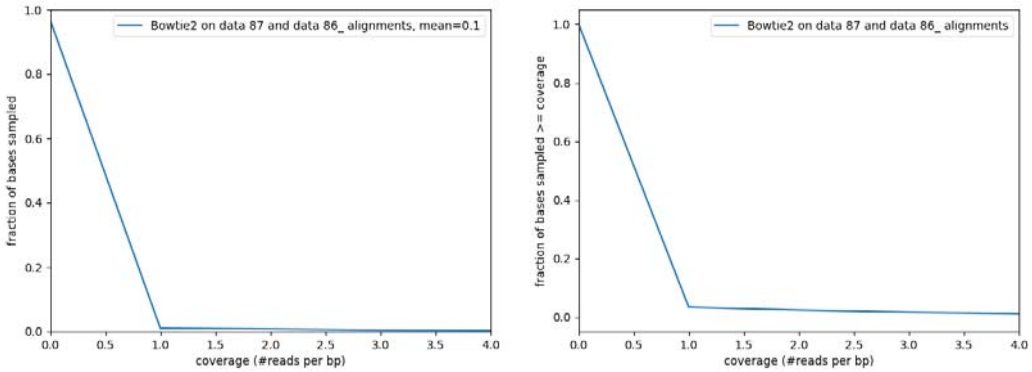


Figure 2: Alignment from reads mapped to (a) *Gallus gallus* and *(b) Taeniopygia guttata*

Gene annotation

Repetitive regions were identified and masked prior to gene prediction. We employed Repeat Modeler(Flynn et al., 2020), a *de novo* transposable element (TE) family identification and modeling package with Repeat Scout tool embedded for identifying the boundaries. The tool works on ascertaining the family relationships from sequence data as it clusters and produces a high quality of TE families (Bourque et al. 2018). We used the Shovill contigs that were generated as an input with 1391 sequences (1696224 bp) for downstream analyses and the resulting library was later checked with *Gallus gallus* repeat libraries. We then used Red (REpeat Detector), a rapid tool for detecting repeats *de novo* on the genomic scale (Girgis et al. 2015). Gene prediction was performed using Augustus (Stanke et al. 2003) with a predefined *Gallus gallus, domesticus* training dataset and softmasking considered for predicting the complete gene models and repeated regions. As we searched for prediction of incomplete genes at the sequence boundaries, we also aimed to predict complete genes. The Red GFF file was used as Augustus produced the fasta files containing predicted coding sequences and that were used for Pfamscan searches (Wang et al. 2013). The resulting Augustus predictions were used as an input with Pfam-A HMM library manually downloaded in stockholm format. We also queried characteristic active site residues if any between the overlaps belonging to the same clans and further checked for functional annotation and domains using InterProScan (Jones et al. 2014) in inferring the Gene Ontology terms (Supplementary Table 3).

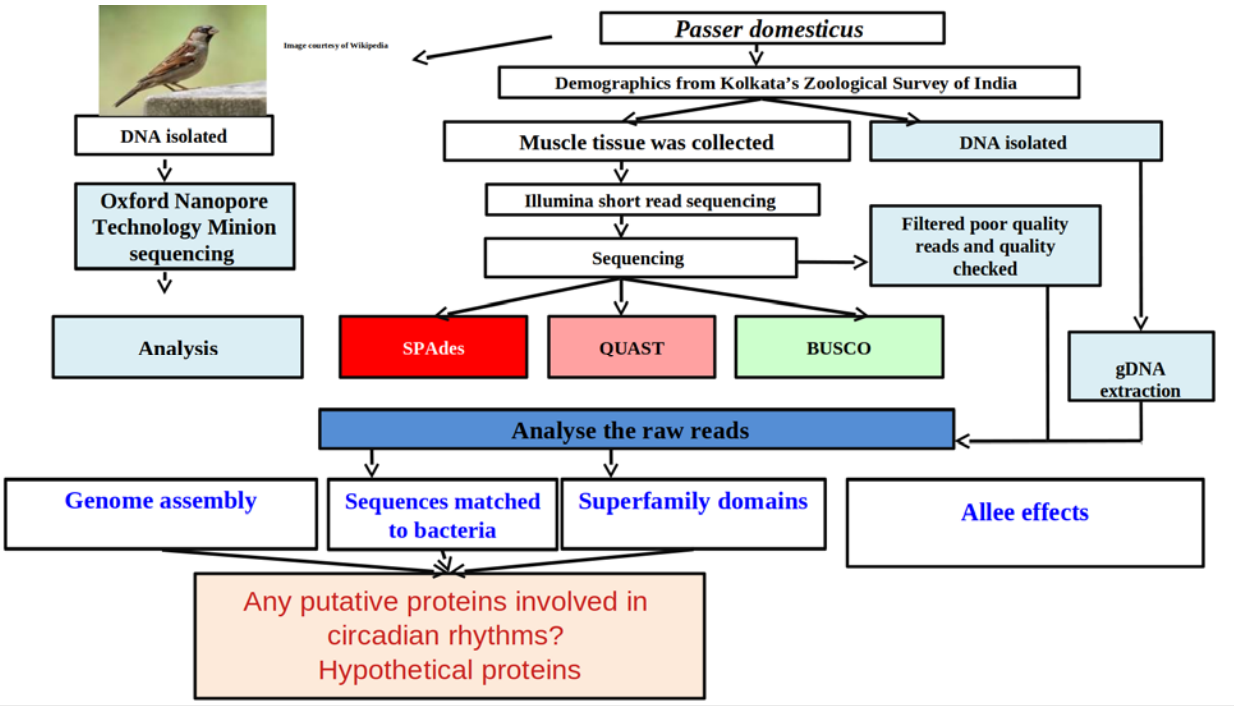


Figure 3: An overview of methodology employed for annotation and assembly of *Passer domesticus*.

Genome comparison and statistics

The *Acanthisitta chloris* genome was obtained from https://db.cngb.org/search/project/CNPhis0000510/ for comparing the contigs with similarity. Batch Entrez, Smith-Watermann algorithm using UVA FASTA from local searches were made for predicting the proteins (https://fasta.bioch.virginia.edu/fasta/fasta_list.html last accessed on October 1, 2023). GffCompare was employed to compare predicted sequences at different levels of granularity, thereby annotating the sequences based on their overlaps or proximity to reference annotation transcripts (Trapnell et al. 2010). Sensitivity and precision metrics were computed with GTF file as an input (Supplementary file) for generating annotated files, yielding the "super-locus" for measuring the accuracy with true positives with other features like true negatives (TN) and false positives (FP) : Sensitivity = TP/(TP+FN) and Precision = TP/(TP+FP). Finally, RefMap and TMAP files (Supplementary Table 4) were obtained measuring the reference transcript that either fully or partially matches a transcript from the GTF and those columns in the file describing the most closely matching reference transcript respectively.

Acknowledgments: PS gratefully acknowledges Department of Forestry, Governments of Rajasthan and India, Zoological Survey of India, and Birla Institute of Scientific Research for providing the apriori permission for collection of bird's sample.

Authors' contributions: PS ideated the project with GC and PBK. VK,SS,GS, DV and PS contributed equally to the work. All the other authors worked towards improvement of tables and figures. PS, GC and PBK proofread the manuscript.

Data Availability: The genome assembly has been deposited at NCBI under BioProject ID: PRJNA1027087

Competing interests: None

Legends to Supplementary Tables and Information

Supplementary Table 1: The final set of contigs obtained for assembly were mapped against NCBI taxa
Supplementary Table 2: The 14947 contigs with 769 of them making 100% query coverage with various hits from taxa
Supplementary Table 3: The Pfam based hits and functional annotation and domains using InterProScan in inferring the Gene Ontology terms
Supplementary Table 4: TMAP files obtained from annotation
Supplementary information: Full GTF file of *Passer domesticus*

References

Adey, A., Kitzman, J. O., Burton, J. N., Daza, R., Kumar, A., Christiansen, L., Ronaghi, M., Amini, S., L. Gunderson, K., Steemers, F. J., & Shendure, J. (2014). In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. Genome Research, 24(12), 2041–2049. https://doi.org/10.1101/gr.178319.114

Botero-Castro, F., Figuet, E., Tilak, M.-K., Nabholz, B., & Galtier, N. (2017). Avian genomes revisited: Hidden genes uncovered and the rates versus traits paradox in birds. Molecular Biology and Evolution, 34(12), 3123–3131. https://doi.org/10.1093/molbev/msx236

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten

things you should know about transposable elements. Genome Biology, 19(1). https://doi.org/10.1186/s13059-018-1577-z

Coombe, L., Zhang, J., Vandervalk, B. P., Chu, J., Jackman, S. D., Birol, I., & Warren, R. L. (2018). arks: Chromosome-scale scaffolding of human genome drafts with linked read kmers. BMC Bioinformatics, 19(1), 234. https://doi.org/10.1186/s12859-018-2243-x

Crick, H.Q.P., Robinson, R.A., Appleton, G F., Clark, N.A., Rickard, A.D. (eds.) (2002) – Investigation into the causes of the decline of starlings and house sparrows in Great Britain – BTO Research Rep. No. 290.

Dadam, D., Robinson, R. A., Clements, A., Peach, W. J., Bennett, M., Rowcliffe, J. M., & Cunningham, A. A. (2019). Avian malaria-mediated population decline of a widespread iconic bird species. Royal Society open science, 6(7), 182197.

De Coster, G., De Laet, J., Vangestel, C., Adriaensen, F., & Lens, L. (2015). Citizen science in action: Evidence for long□term, region wide house sparrow declines in Flanders, Belgium. Landscape and Urban Planning, 134,139–146.https://doi.org/10.1016/j.Landurbplan.2014.10.020.

Dhar, R., Seethy, A., Pethusamy, K., Singh, S., Rohil, V., Purkayastha, K., Mukharjee, I., Goswami, S., Singh, R., Raj, A., Srivastava, T., Acharya, S., Rajashekhar, B., Karmakar, S. (2019). De novo assembly of the Indian blue peacock (Pavo cristatus) genome using Oxford Nanopore technology and Illumina sequencing, GigaScience, Volume 8, Issue 5, May 2019, giz038, https://doi.org/10.1093/gigascience/giz038

Ducrest, A. L., Neuenschwander, S., Schmid□Siegert, E., Pagni, M., Train, C., Dylus, D., ... & Dessimoz, C. (2020). New genome assembly of the barn owl (Tyto alba alba). Ecology and Evolution, 10, 2284-2298.

Dunning Hotopp JC. Horizontal gene transfer between bacteria and animals. Trends Genet. 2011 Apr;27(4):157-63. doi: 10.1016/j.tig.2011.01.005. Epub 2011 Feb 18.

Enbody ED, Sendell-Price AT, Sprehn CG, Rubin CJ, Visscher PM, Grant BR, Grant PR, Andersson L. Community-wide genome sequencing reveals 30 years of Darwin's finch evolution. Science. 2023 Sep 29;381(6665):eadf6218. doi: 10.1126/science.adf6218.

Feng, S., J. Stiller, Y. Deng, J. Armstrong, et al. (2020). Dense sampling of bird diversity increases the power of comparative genomics. Nature, 857:252-257.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences, 117(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Frankl-Vilches C , et al. 2015. Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. Genome Biol. 16(1):19. DOI 10.1186/s13059-014-0578-9.

Friis, G., Fandos, G., Zellmer, A. J., McCormack, J. E., Faircloth, B. C., & Milá, B. (2018). Genomewide signals of drift and local adaptation during rapid lineage divergence in a songbird. Molecular Ecology, 27, 5137-5153.

Gascoigne, J., " A theoretical and experimental study of Allee effects" (2003). Dissertations, Theses, and Masters Projects. William & Mary. Paper 1539616659. https://dx.doi.org/doi:10.25773/v5-qwvk-b742.

Girgis, H. Z. (2015). Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinformatics, 16(1). https://doi.org/10.1186/s12859-015-0654-5

Gulati V (2005): House sparrow on verge of extinction. Tribune News Service. Chandigarh, India. Retrieved September 20, 2005, from http://www.Tribuneindia.com/2005/20050920/cth1.htm#11

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ... & Yuri, T. (2008). A phylogenomic study of birds reveals their evolutionary history. science, 320(5884), 1763-1768. DOI: 10.1126/science.1157704

Illera, J.C., Rando, J.C., Rodriguez-Exposito, E., Hernández, M., Claramunt, S. and Martín, A. (2018), Acoustic, genetic, and morphological analyses of the Canarian common chaffinch complex Fringilla coelebs ssp. reveals cryptic diversification. J Avian Biol, 49: https://doi.org/10.1111/jav.01885

Jackman, S. D., Coombe, L., Chu, J., Warren, R. L., Vandervalk, B. P., Yeo, S., Xue, Z., Mohamadi, H., Bohlmann, J., Jones, S. J. M., & Birol, I. (2018). tigmint: Correcting assembly errors using linked reads from large molecules. BMC Bioinformatics, 19(1), 393. https://doi.org/10.1186/s12859-018-2425-6

Jarvis, E.D., Mirarab, S., Aberer, A., Hounde, P., Li, C., Ho, S., et al. (2015). Phylogenomics analyses data of the avian phylogenomics project. GigaScienceDatabase. http://dx.doi.org/10.5524/101041

Jarvis, E.D., Mirarab, S., Aberer, A., Hounde, P., Li, C., Ho, S., et al. (2014). Whole Genome Analyses Resolve the Early Branches to the Tree of Life of Modern Birds.Science, 346(6215), 1320-1331 DOI:10.1126/science.1253451

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification.

Bioinformatics. 2014 May 1;30(9):1236-40. doi: 10.1093/bioinformatics/btu031. Epub 2014 Jan 21. PMID: 24451626; PMCID: PMC3998142.

Kraus, R. H., & Wink, M. (2015). Avian genomics: fledging into the wild!. Journal of Ornithology, 156, 851-865.

Louha, S., Ray, D. A., Winker, K., & Glenn, T. (2019). A high-quality genome assembly of the North American Song Sparrow, Melospiza melodia. bioRxiv, 850990.

Magallanes-Alba ME, Baricalla A, Rego N, Brun A, Karasov WH, Caviedes-Vidal E. Rapid genome functional annotation pipeline anchored to the house sparrow (Passer domesticus, Linnaeus 1758) genome reannotation. Biol Methods Protoc. 2023 Jul 7;8(1):bpad013. doi: 10.1093/biomethods/bpad013. PMID: 37521110; PMCID: PMC10373626.

María Recuerda, Joel Vizueta, Cristian Cuevas-Caballé, Guillermo Blanco, Julio Rozas, Borja Milá, Chromosome-Level Genome Assembly of the Common Chaffinch (Aves: Fringilla coelebs): A Valuable Resource for Evolutionary Biology, Genome Biology and Evolution, Volume 13, Issue 4, April 2021, evab034, https://doi.org/10.1093/gbe/evab034

Miller, M. M., & Taylor, R. L. Jr (2016). Brief review of the chicken Major Histocompatibility Complex: The genes, their distribution on chromosome 16, and their contributions to disease resistance. Poultry Science, 95(2), 375–392. https://doi.org/10.3382/ps/pev379

Mohring, B., Henry, PY., Jiguet, F. et al. Investigating temporal and spatial correlates of the sharp decline of an urban exploiter bird in a large European city. Urban Ecosyst 24, 501–513 (2021). https://doi.org/10.1007/s11252-020-01052-9.

Ou, S., Chen, J., & Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Research, 46(21), e126. https://doi.org/10.1093/nar/gky730

Peñalba, J. V., Deng, Y., Fang, Q., Joseph, L., Moritz, C., & Cockburn, A. (2019). Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (Malurus cyaneus). Molecular Ecology Resour. https://doi.org/10.1111/1755-0998.13124

Peona V, Blom MPK, Xu L, et al. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. Mol Ecol Resour. 2021;21:263–286. https://doi.org/10.1111/1755-0998.13252

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., ... & Wolf, J. B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science, 344, 1410-1414.

Rahmani Asad R (2008): Flight to extinction. Spectrum,The Tribune. Retrieved June 8, 2008, from

Sang, H. (2004). Prospects for transgenesis in the chick. Mech. Dev. 121, 1179–1186. https://doi.org/10.1016/j.mod.2004.05.012

Seemann, T. (2017). Shovill: faster SPAdes assembly of Illumina reads. Tilgjengelig fra Tilgjengelig fra. https://github.com/tseemann/shovill

Sendell-Price AT, Ruegg KC, Clegg SM. Rapid morphological divergence following a human-mediated introduction: the role of drift and directional selection. Heredity (Edinb). 2020 Apr;124(4):535-549. doi: 10.1038/s41437-020-0298-8. Epub 2020 Feb 20.

Seppey, M., Manni, M., Zdobnov, E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar, M. (eds) Gene Prediction. Methods in Molecular Biology, vol 1962. Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9173-0_14

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics, 19(Suppl 2), ii215â-“ii225. https://doi.org/10.1093/bioinformatics/btg1080

Stoltzfus A, Norris RW. On the Causes of Evolutionary Transition:Transversion Bias. Mol Biol Evol. 2016 Mar;33(3):595-602. doi: 10.1093/molbev/msv274. Epub 2015 Nov 25. PMID: 26609078; PMCID: PMC7107541.

Storch V, Welsch U, Wink M (2013) Evolutionsbiologie (in German), vol 3. Spektrum Springer, Heidelberg. https://link.springer.com/book/10.1007/978-3-642-32836-7

Summers-Smith, J.D.. (2003). The decline of the House Sparrow: A review. British Birds. 96. 439-446.

Suravajhala P and Acharya K. WHITHER GENOME SEQUENCING OF PASSER? ECB. 2014; 3(11): 1086-1087

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, 28(5), 511â515. https://doi.org/10.1038/nbt.1621

Wang W, Wang F, Hao R, Wang A, Sharshov K, Druzyaka A, Lancuo Z, Shi Y, Feng S. (2020). First de novo whole genome sequencing and assembly of the bar-headed goose. PeerJ 8:e8914. https://doi.org/10.7717/peerj.8914.

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., & Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Research, 41(6), e74–e74. https://doi.org/10.1093/nar/gkt006

Warren, W. C., Hillier, L. D. W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K. D., Thibaud-Nissen, F., Schneider, V., Mansour, T. A., Brown, C. T., Zimin, A., Hawken, R., Abrahamsen, M., Pyrkosz, A. B., Morisson, M., Fillon, V., … Cheng, H. H. (2017). A new chicken genome assembly provides insight into avian genome structure. G3: Genes, Genomes, Genetics, 7(1), 109–117. https://doi.org/10.1534/g3.116.035923

Watson, M., & Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. Nature Biotechnology, 37(2), 124–126. https://doi.org/10.1038/s41587-018-0004-z

Yeo, S., Coombe, L., Warren, R. L., Chu, J., & Birol, I. (2017). arcs: Scaffolding genome drafts with linked reads. Bioinformatics, 34(5), 725–731. https://doi.org/10.1093/bioinformatics/btx675

Zhang G et al (2014b) Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311 1320. https://doi:10.1126/science.1251385

Zhang, G. Bird sequencing project takes off. Nature 522, 34 (2015). https://doi.org/10.1038/522034d.

Zhang, G., Parker, P., Li, B., Li, H., & Wang, J. (2012). The genome of Darwin’s finch (Geospiza fortis). GigaScience. Database.

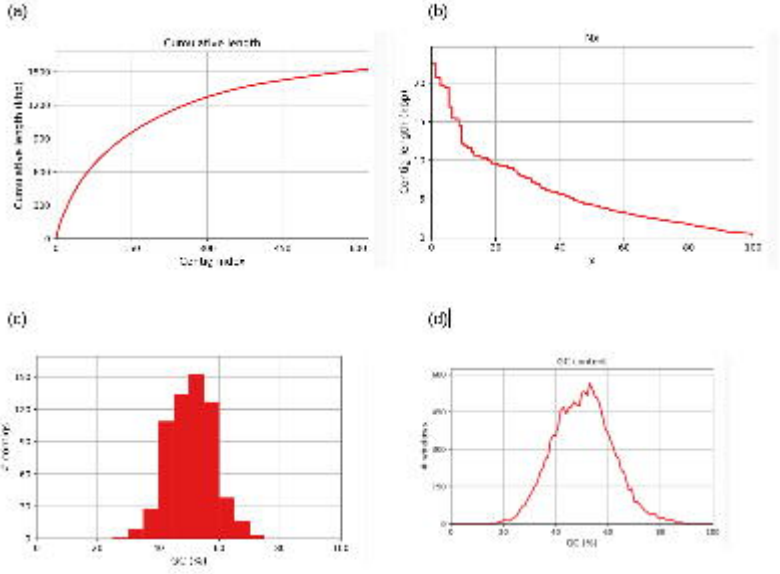


Figure 1: (a) Cumulative length vs contig length with (b) N50 (Nx) plotted, (c) GC% against number of contigs and (d) GC content against number of windows

(a) (b)

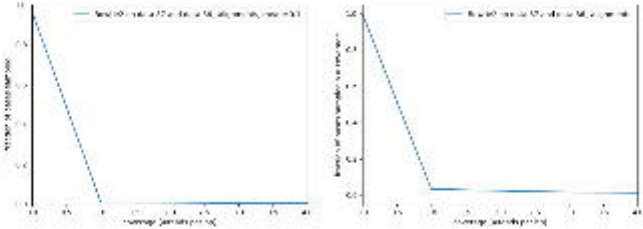
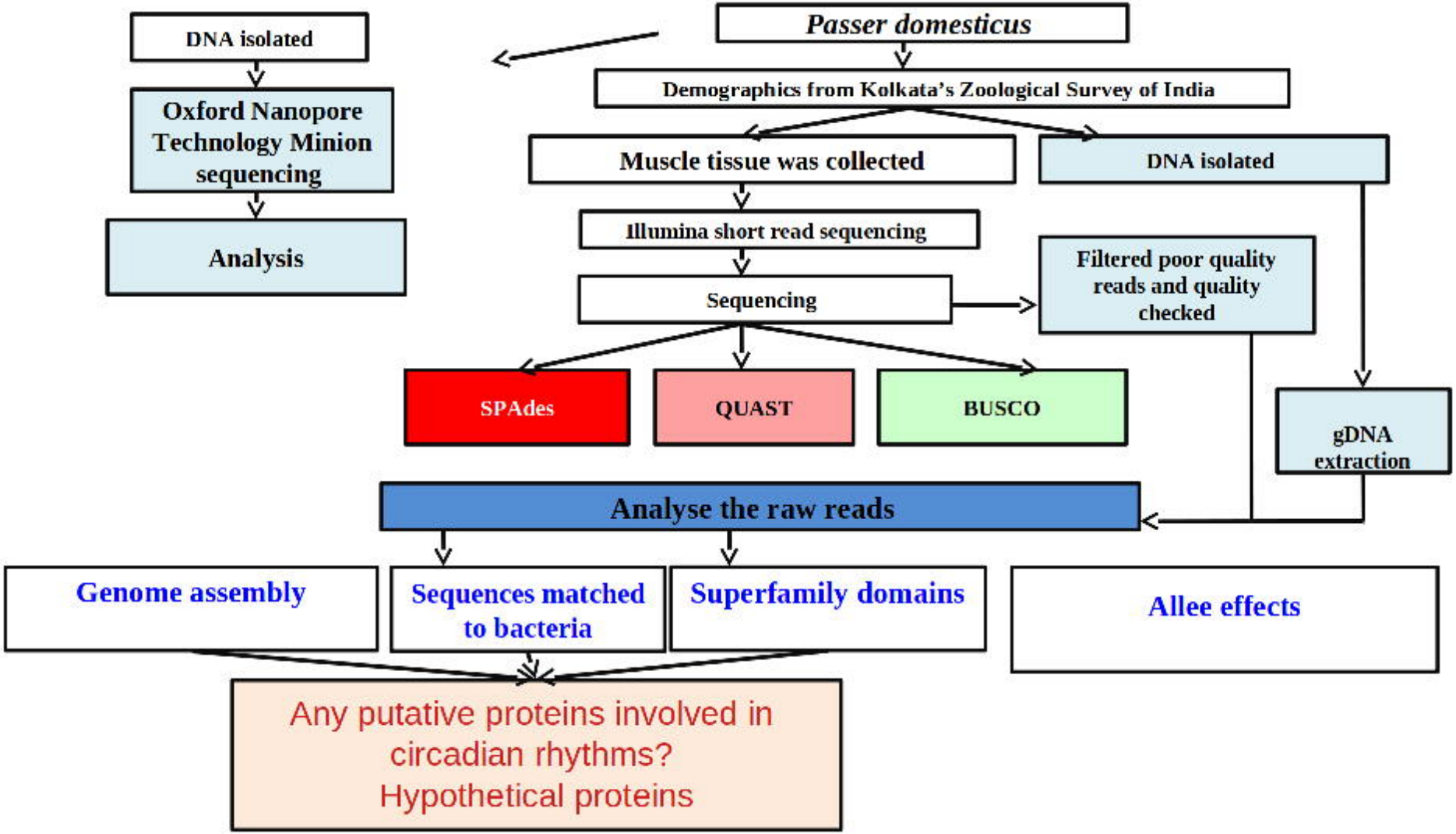


Figure 2: Alignment from reads mapped to (a) *Gallus gallus* and *(b) Taeniopygia guttata*


Passer domesticus
DNA isolated
Oxford Nanopore Technology Minion sequencing
Analysis
Demographics from Kolkata's Zoological Survey of India
Muscle tissue was collected
DNA isolated
Illumina short read sequencing
Sequencing
Filtered poor quality reads and quality checked
SPAdes
QUAST
BUSCO
gDNA extraction
Analyse the raw reads
Genome assembly
Sequences matched to bacteria
Superfamily domains
Allee effects
Any putative proteins involved in circadian rhythms?
Hypothetical proteins