DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„ Conservation of the Little Vermillion Flycatcher population in Santa Cruz and genetic assessment for future reintroduction planning to reestablish former range“

verfasst von / submitted by

David Jhon Anchundia Gonzalez

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2024/ Vienna 2024

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 794 685 437
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Biologie
Betreut von / Supervisor:	Mag. Dr. Sabine Tebbich, Privatdoz.
Betreut von / Supervisor:	Mag. Dr. Birgit Fessl

ACKNOWLEDGEMENTS

I dedicate this work to my family and friends for their unwavering support during difficult times. A special dedication goes to the memory of my beloved mother, Juana, whose inspiration and encouragement to explore and dedicate my career to science have been a guiding light in my life. I also want to express my heartfelt gratitude to my dear Courtney for her constant support throughout this journey. It has been incredibly enriching to navigate this path together as two PhD students at the University of Vienna.

I am deeply thankful to all those who supported me and helped me grow during this period of learning and education. First, I would like to thank Dr. John Dumbacher, for being a source of inspiration and for encouraging my interest in genetic studies. Before embarking on my PhD, I had limited knowledge of genetics. Thanks to John and his team, I was able to immerse myself in this field, and this foundation will be crucial for the development of my career after my PhD. I would also like to extend my gratitude to Dr. Sabine Tebbich, my PhD advisor, for her invaluable supervision, scientific guidance, and financial support throughout my time at the University of Vienna. Without her support, this study would not have been possible. Additionally, I want to thank everyone who contributed to this study, particularly those who assisted with data collection and provided support. Special thanks to Dr. Birgit Fessl, Heinz Richner, Agustin Gutierrez, Peter Pibaque, Dr. Charlotte Causton, and my sincere gratitude to Novarino Castillo and Kleber Abad, who played a vital role in making the study possible by clearing the study site for data collection. I would like to thank the park rangers of the Galápagos National Park for their assistance with fieldwork and logistics, especially Rafael Chango and Christian Sevilla. To everyone who was part of this PhD journey, I offer my deepest thanks for your support.

Also, I thank the World Wildlife Fund, Mr. Russell E. Train and Education for Nature, Agreement #RH94 for the scholarship they awarded me to pursue my PhD. I acknowledge the Lakeside Fellowship from the California Academy of Sciences and a donation from Michael Simon to the California Academy of Sciences for the development of my study. I am grateful to other funding sources such as the Galapagos Invasive Species Fund; Galapagos Conservation Trust; Galapagos Conservancy; Swiss Association of Friends of the Galapagos Islands; COmON Foundation; Re:wild; Lindblad Expeditions - National Geographic; and Peter and Kris Norvig. Grant awarded to Sabine

Tebbich and Rebecca Hood-Nowotny by the National Geographic Science Fund. Austria (grant number P 32555-B). This publication is contribution number 2696 of the Charles Darwin Foundation for the Galapagos Islands. The data collected for this study were approved by the Galapagos National Park Directory PC-30-19, PC-11-20, PC-06-21, PC-22-23 and the genetic permit approved by the Ministry of the Environment of Ecuador MAE-DNB-CM-2016-0043.

Table of Contents

General Introduction

Species conservation

Conservation of species diversity is a priority for maintaining the balance, stability and resilience of ecosystems (Ives and Carpenter, 2007; Hong et al., 2022). The diversity of species and their benefits in an ecosystem have been widely studied because each species can have multiple functions in its niche, such as in the energy flow in the food web, pollination, nutrient cycling, etc. (Lyons et al., 2005). Unfortunately, in recent decades there has been a global crisis of species loss and extinction occurring at an alarming rate (Sigwart et al., 2018; Driscoll et al., 2018). Even remote sites, such as the Galapagos Islands, are not safeguarded from this biodiversity loss (Grant et al., 2005; Benitez-Capistros et al., 2014; Dvorak et al., 2020). Other well-known archipelagos, including Hawaii and other isolated islands, have recently suffered a significant loss of fauna (Paulay, 1994; Boyer, 2008). Terrestrial vertebrate species on islands, especially birds, are the groups declining the most (Steadman, 1995; Alcover et al., 1998; Blackburn et al., 2004) and are presently in danger of island extirpation. Therefore, identifying key drivers for these changes in the Galapagos Islands and elsewhere is necessary.

An important key factor to prevent species loss is the preservation of food resources. Resource availability and proper vegetation structure are essential for greater richness and diversity of bird populations in an ecosystem (Ferger et al., 2014; Wolfe et al., 2014). However, habitat alteration due to anthropogenic effects can have detrimental effects on birds (Callaghan et al., 2023). For example, consequences of agricultural practices include habitat alteration and habitat loss, which are considered the main causes of declines and extinctions of bird populations (Green et al., 2005; Stanton et al., 2018). However, some species are more resistant to habitat alterations than others (Guénette and Villard, 2005). Another challenge in preventing species loss is the impact of invasive plants, which alter soil properties and microbial taxa, producing changes in nutrient cycling in an ecosystem (Hawkes et al., 2005; Gibbons et al., 2017). Therefore, these invasive species can change the vegetation structure, altering species diversity, species interactions, and primary production (Grice, 2006). Furthermore, invasive plants can cause changes in species composition and cause declines in species richness, specifically of arthropods, in an ecosystem (Litt et al., 2014), which in turn, affects the survival of birds.

Studying population dynamics allows us to identify the greatest threats and identify which life stage is most affected. A simple model of population dynamics for a population to grow or remain stable is influenced by demographic processes: births, survival, migration and immigration (Keeler and Joern, 1995; White, 2000; Lima and Jaksic, 1999). Each of these processes is influenced by

abiotic and biotic factors (Andrewartha and Birch, 1954; Hixon et al., 2002). For example, climate change, an abiotic factor, could alter seasonal reproduction and survival (Jiguet et al., 2010), while biotic factors, including parasitism and food availability, can affect population dynamics of a species (Loye, 1995; Lafferty and Kuris, 1999). Therefore, negative effects on these demographic processes might lead to population declines and can affect the long-term survival of a population.

Furthermore, theories on population genetics propose that if a species' population is drastically reduced, it will no longer have a sufficient population size to generate and accumulate mutations and adapt to sudden changes in its environment (Loeschcke et al., 2013). Studies on threatened populations show that small populations have a risk of disappearing due to genetic bottlenecks that entail loss of heterozygosity, shifts in allele frequencies, reduced genetic diversity, inbreeding and genetic drift (Kirkpatrick and Jarne, 2000; Spigler et al., 2017). A review study of 53 species of birds found that population reductions had detrimental effects on their reproduction (Heber and Briskie, 2010). This occurs when a small population is exposed to stochastic environmental events, as well as facing genetic problems (Lande, 1993; Keller and Waller, 2002). For example, infertility or congenital problems might arise from mating with genetically closely related individuals (Jamieson and Ryan, 2000; Briskie and Machiston, 2004). Therefore, the loss of genetic diversity could have an impact on reproduction and the viability of the population. Additionally, threatened and endangered bird species are often observed to have a biased sex ratio, usually consisting of more males than females, which poses additional risks, lowers the genetically-effective population size, and hampers population recovery (Robinson et al., 2014; Donald, 2007).

Study Site

The Galapagos is a remote archipelago located in the Pacific Ocean with high species endemism both within and among islands (Emerson, 2002). Of the 28 endemic landbird species in Galapagos, 16 species are in decline and face conservation challenges (BirdLife International, 2024). All 16 species in decline have something in common - they all depend primarily on an arthropod diet to survive. The decline of arthropods and the decline in population sizes of insectivorous birds is a current global pattern (BirdLife International, 2022; Tallamy and Shriver, 2021; Sekercioglu, 2002; Forister et al., 2019), which is also documented in the Galapagos (Grant et al., 2005; Dvorak et al., 2012, 2017, 2020). If this trend continues in the coming decades, the number of extinctions will increase in Galapagos. However, we still lack basic knowledge of the taxa and populations of birds in Galapagos despite substantial research conducted in the last century. In fact, certain groups that have been poorly studied are in decline; therefore it is likely that unique populations could be lost and their value or role in the ecosystem may also be lost.

Focal Study Species

Here, I present a series of studies on the vermilion flycatchers in the Galapagos Islands, Ecuador. This bird group is one of the most seriously impacted across the archipelago and lacking important data for its conservation. One species endemic to San Cristobal Island, the Least Vermilion Flycatcher (*Pyrocephalus dubius*), has recently been declared extinct (Dvorak et al., 2020), which may be the first vertebrate species extinction in the Galapagos in modern times. The other recognized vermilion flycatcher species in Galapagos, the Little Vermilion Flycatcher (*Pyrocephalus nanus*), is facing drastic population declines on multiple islands (BirdLife International, 2024) and is currently considered 'near threatened' (BirdLife International, 2024). A century ago, its populations were present on 12 islands in the Galapagos archipelago (Gifford, 1919; Swarth, 1931). However, vermilion flycatchers are no longer present on four islands and are on the brink of disappearing from three more islands (Fessl et al., 2017). Furthermore, some of these extirpated *P. nanus* populations were morphologically distinct and were possibly unique taxonomic lineages, endemic island species, or subspecies (Ridgway, 1897; Snodgrass and Heller, 1904; Swarth, 1931).

Pyrocephalus demonstrates a wide distribution both geographically and ecologically and faces challenges that many insectivorous birds have in common in Galapagos. Therefore, this genus is an excellent choice for studying general principles in island conservation, avian ecology, island biogeography, and other fields relevant in the Galapagos archipelago today. In my PhD thesis, I focus mainly on the Little Vermilion Flycatcher (*P. nanus*), since *P. dubius* has not been documented since 2008 and is considered extinct. Santa Cruz Island has a particularly vulnerable population of *P. nanus*, which is on the brink of disappearing. As recently as two decades ago, the population was much larger, but it has more recently been reduced to fewer than 40 territories (Fessl et al., 2017; Leuba et al., 2020).

Population dynamics of *P. nanus* in Santa Cruz has not been studied in depth, but several factors might be affecting its demography. Reproduction in particular appears to be extremely low on Santa Cruz and Isabela islands (Leuba et al., 2020; Mosquera et al., 2022). Therefore, for recovery of this population, it will be critical to understand why reproduction and recruitment are so low. One major biotic factor affecting this population is the introduction of invasive species (Fessl et al., 2018; Rentería et al., 2021), which impact the reproduction and survival of this species. One invasive species in the Galapagos is the Avian Vampire Fly (*Philornis downsi*), which parasitizes nestlings (Leuba et al., 2020; Mosquera et al., 2022). This fly is native to South America and Trinidad (Bulgarella et al., 2015; Koop et al., 2021) and was introduced to Galapagos, with its first record in an entomological collection in 1964 (Causton et al., 2006). To complete its life cycle, the Avian Vampire Fly needs an avian host during its larval stages, during which the larvae consume blood and tissue, mainly from

nestlings (Fessl et al., 2018). This significantly increases nestling mortality and reduces recruitment of new individuals (Leuba et al., 2020; Mosquera et al., 2022).

Furthermore, invasive plants are another key biotic factor that may influence *P. nanus* reproduction and recruitment due to their negative impact in the ecosystems in Santa Cruz (Leuba et al., 2020), such as that by the invasive blackberry (*Rubus niveus*). This invasive species has rapidly expanded in habitat areas of *P. nanus*, reducing the diversity of native and endemic plants and changing the structure of the understory (Rentería et al., 2012; Walentowitz et al., 2021). The blackberry plants alter invertebrate availability and can affect foraging behavior of endemic landbirds, including insectivores, leading to lower breeding success for Galapagos landbirds (Cimadom et al., 2014; Jäger et al., 2015; Hood-Nowotny et al., 2023). On pristine areas *P. nanus* predominantly catches arthropods near the ground from low perches (Anchundia pers obs.). However, in disturbed island ecosystems the blackberry forms a dense understory leaving few open areas near the ground for *P. nanus* to hunt for arthropods (Leuba et al., 2020). Therefore, these stressors must be removed to facilitate successful foraging, thus potentially positively influencing reproductive success and population growth.

Additionally, theories of population genetics suggest that inbreeding could potentially distort sex ratio (Frankham and Wilcken, 2006). The small population size of *P. nanus* in Santa Cruz may lead to inbreeding and other forms of genetic drift that reduce genetic diversity and may threaten this population. Given the population's biased sex ratio, it may already be experiencing inbreeding. We observed at least twice as many males as females in the Santa Cruz population. Furthermore, it is necessary to evaluate whether a genetic rescue, through translocations from other islands to increase genetic diversity, is needed for this population in Santa Cruz. To determine this, it is necessary to conduct a genetic evaluation on populations of *P. nanus* from all islands of the archipelago to determine which population is more closely related genetically to the population on Santa Cruz. Moreover, it is unknown if there is migration between each island or if each population of *P. nanus* is a different lineage. There is a perception that certain populations could be very distinct since morphological differences have been seen using museum collections, and some authors have even divided these populations into multiple species (up to five). Therefore, it is important to elucidate the evolutionary history of this group in Galapagos.

Aims of This PhD Thesis

In my PhD thesis, I aimed to increase the knowledge about *P. nanus* in Galapagos to inform conservation and reintroduction initiatives. Firstly, I aimed to determine how habitat restoration and removal of invasive species influences the reproductive success of *P. nanus* in Santa Cruz. Secondly, I

aimed to evaluate the genetic diversity and heterozygosity of the small population in Santa Cruz and its long-term viability. Finally, I analyzed the taxonomy of the *Pyrocephalus* genus in Galapagos to determine if the populations on each island have distinct genetic differences and/or if there is frequent genetic exchange.

Given this, I aimed to address the following questions in my PhD thesis:

I. Does habitat restoration improve breeding success of *P. nanus* in Santa Cruz Island?

II. What is the current genetic diversity of *P. nanus* in Santa Cruz Island and implications of this for its conservation, utilizing the whole genome?

III. What is the level of genetic isolation between populations of *P. nanus* from different islands in the Galapagos?

To answer these questions, I conducted controlled field experiments involving invasive species removal to study bird behaviour and reproduction, conducted whole genome sequencing, and performed bioinformatics.

In **Chapter 1**, I aimed to investigate whether habitat management can improve the reproductive success of *P. nanus* to prevent its extirpation from Santa Cruz Island. I hypothesized that the poor reproductive success is related to the inaccessibility of food produced by the dense understory formed by blackberry and the parasitism of *P. downsi*. To improve habitat quality and enhance the reproductive success of *P. nanus* in Santa Cruz, in seven 1 ha areas invasive blackberry was removed to provide more open areas to facilitate *P. nanus* in foraging. To measure the impact of management, I compared foraging and breeding success of *P. nanus* in managed areas to areas where habitat restoration was not carried out. In managed and unmanaged areas, I compared data on their foraging behavior, feeding rate during incubation, rearing of nestlings, foraging success, pair formation, nesting attempts, and reproductive success. I predicted that habitat recovery will give these birds more opportunities to find more food close to the ground and will aid in improving their reproductive success. I discussed the results from this study in relation to using habitat management as a conservation tool.

In **Chapter 2**, I conducted an in-depth study of *P. nanus* using molecular techniques. I sequenced and assembled the whole genome of this species using the De Novo genome assembly method to use this as a reference genome. On this reference genome, we identified sexual chromosomes and thousands of genes of interest. This genome is ideal for identifying single nucleotide polymorphism (SNPs) that are under selection and for performing deeper genetic studies of

population structure and evolutionary history of *Pyrocephalus*. I conducted this further research in the next chapter.

In **Chapter 3**, I investigated the current level of genetic diversity of the small population of *P. nanus* on Santa Cruz Island and its heterozygosity, comparing it with the genetic diversity of birds that lived 115 and 60 years ago in Santa Cruz, which are stored in the museum collections at the California Academy of Sciences. The museum specimens used for this study were collected at different sites on Santa Cruz Island, when *P. nanus* was still abundant. This allowed us to develop a baseline of how the genetic diversity of this population was in the past. For comparison of individuals from the past and present, I sequenced samples from these birds to obtain high coverage genomes and aligned them to the reference genome to obtain a large dataset and high-resolution results. Due to several characteristics that we currently observe in *P. nanus* in Santa Cruz, I predicted that the genetic diversity of this population has declined, and therefore, the population may have a high level of inbreeding.

Additionally, in this Chapter, I focus to assess the level of genetic isolation between populations of *P. nanus* from different islands in the Galapagos. Several authors in the early 20th century suggested that populations of *P. nanus* on certain islands in the Galapagos differ morphologically; however, this has not been evaluated in depth from a genetic perspective. Due to the small population size of *P. nanus* in Santa Cruz, it is necessary to know if this population is genetically unique or if it shares similarities with other islands. This information is pertinent to help inform how to carry out translocations, if necessary. In this chapter, I aimed to understand the level of isolation versus connectedness of different *Pyrocephalus* populations in Galapagos using whole genome sequences, mapped to a reference genome. For islands where *Pyrocephalus* populations are already considered extinct or extirpated, I took samples from museum skins from the California Academy of Sciences to obtain DNA and sequences. The objectives were to 1) build a phylogeny of all *Pyrocephalus* populations in Galapagos and evaluate if there are distinct genetic groups, 2) compare their morphology and acoustics between islands, and 3) review their taxonomy, as necessary. The information from this study can be used in future rewilding (reintroduction) programs currently being implemented in Galapagos, with the aim to recover the original distribution of species that have disappeared through translocation, using individuals from the most genetically similar population.

Overall, the information gained from my PhD is crucial for supporting bird conservation efforts in the Galapagos, not only for the *Pyrocephalus* group but also for other species in decline. My research conducted on habitat restoration and control of invasive species could be applied effectively in Galapagos or elsewhere. Also, the integration of new molecular techniques into conservation studies is gradually increasing in Galapagos, and my research will further encourage local researchers to develop new projects aimed at conserving the unique species in the archipelago.

REFERENCES

Alcover, J.A., Sans, A. and Palmer, M. (1998). The extent of extinctions of mammals on islands. *Journal of Biogeography*, 25(5), 913-918.

Andrewartha, H.G. and Birch, L.C. (1954). *The distribution and abundance of animals*. No. Edn 1.

Benitez-Capistros, F., Hugé, J. and Koedam, N. (2014). Environmental impacts on the Galapagos Islands: Identification of interactions, perceptions and steps ahead. *Ecological Indicators*, 38, 113-123.

BirdLife International. (2022) State of the World's Birds 2022: Insights and solutions for the biodiversity crisis. Cambridge, UK: BirdLife International.

BirdLife International. (2024). *Pyrocephalus nanus*. The IUCN Red List of Threatened Species 2023: e.T103682926A172654604. https://dx.doi.org/10.2305/IUCN.UK.2023-1.RLTS.T103682926A172654604.en. Accessed on 29 March 2024.

Blackburn, T.M., Cassey, P., Duncan, R.P., Evans, K.L. and Gaston, K.J. (2004). Avian extinction and mammalian introductions on oceanic islands. *Science*, 305(5692), 1955-1958.

Briskie, J.V., and Mackintosh, M. (2004). Hatching failure increases with severity of population bottlenecks in birds. *Proceedings of the National Academy of Sciences USA*, 101(2), 558-561.

Boyer, A.G. (2008). Extinction patterns in the avifauna of the Hawaiian islands. *Diversity and Distributions*, 14(3), 509-517.

Bulgarella, M., Quiroga, M.A., Brito Vera, G.A., Dregni, J.S., Cunninghame, F., Mosquera Muñoz, D.A., Monje, L.D., Causton, C.E. and Heimpel, G.E. (2015). *Philornis downsi* (Diptera: Muscidae), an avian nest parasite invasive to the Galapagos Islands, in mainland Ecuador. *Annals of the Entomological Society of America*, 108(3), 242-250.

Callaghan, C.T., Chase, J.M. and McGlinn, D.J. (2023). Anthropogenic habitat modification causes nonlinear multiscale bird diversity declines. *Ecography*, e06759.

Causton, C.E., Peck, S.B., Sinclair, B.J., Roque-Albelo, L., Hodgson, C.J. and Landry, B. (2006). Alien insects: threats and implications for conservation of Galápagos Islands. *Annals of the Entomological Society of America*, 99(1), 121-143.

Cimadom, A., Ulloa, A., Meidl, P., Zöttl, M., Zöttl, E., Fessl, B., Nemeth, E., Dvorak, M., Cunninghame, F., and Tebbich, S. (2014). Invasive parasites, habitat change and heavy rainfall reduce breeding success in Darwin's finches. *PloS One*, *9*, e107518.

Donald, P.F. (2007). Adult sex ratios in wild bird populations. *Ibis*, 149(4), 671-692.

Driscoll, D.A., Bland, L.M., Bryan, B.A., Newsome, T.M., Nicholson, E., Ritchie, E.G. and Doherty, T.S. (2018). A biodiversity-crisis hierarchy to evaluate and refine conservation indicators. *Nature Ecology & Evolution*, 2(5), 775-781.

Dvorak, M., Fessl, B., Nemeth, E., Anchundia, D., Cotin, J., Schulze, C.H., Tapia, W. and Wendelin, B. (2020). Survival and extinction of breeding landbirds on San Cristóbal, a highly degraded island in the Galápagos. *Bird Conservation International*, 30(3), 381-395.

Dvorak, M., Fessl, B., Nemeth, E., Kleindorfer, S. and Tebbich, S. (2012) Distribution and abundance of Darwin's finches and other land birds on Santa Cruz Island, Galapagos: evidence for declining populations. *Oryx*, 46, 78-86.

Dvorak, M., Nemeth, E., Wendelin, B., Herrera, P., Mosquera, D., Anchundia, D., Sevilla, C., Tebbich, S. and Fessl, B. (2017). Conservation status of landbirds on Floreana: the smallest inhabited Galapagos Island. *Journal Field Ornithology*, 88(2), 132-145.

Emerson, B.C. (2002). Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology*, 11(6), 951-966.

Ferger, S.W., Schleuning, M., Hemp, A., Howell, K.M. and Böhning-Gaese, K. (2014). Food resources and vegetation structure mediate climatic effects on species richness of birds. *Global Ecology and Biogeography*, 23(5), 541-549.

Fessl, B., Anchundia, D., Carrion, J., Cimadom, A., Cotín, J., Cunninghame, F., Dvorak, M., Mosquera, D., Nemeth, E., Sevilla, C., Tebbich, S., and Causton, C. (2017). Galapagos landbirds (passerines, cuckoos, and doves): Status, threats and knowledge gaps. In Galapagos Reports 2015-2016. GNPS, GCREG, CDF and GC. Puerto Ayora, Galapagos, Ecuador.

Fessl, B., Heimpel, G. and Causton, C. (2018). Invasion of an avian nest parasite, *Philornis downsi*, to the Galapagos Islands: colonization history, adaptations to novel ecosystems, and conservation challenges. *Disease ecology: Galapagos birds and their parasites*, 213-266.

Forister, M.L., Pelton, E.M. and Black, S.H. (2019). Declines in insect abundance and diversity: We know enough to act now. *Conservation Science and Practice*, *1*(8), e80.

Frankham, R. and Wilcken, J. (2006). Does inbreeding distort sex-ratios?. *Conservation Genetics*, *7*(6), 879-893.

Gibbons, S.M., Lekberg, Y., Mummey, D.L., Sangwan, N., Ramsey, P.W. and Gilbert, J.A. (2017). Invasive plants rapidly reshape soil properties in a grassland ecosystem. *MSystems*, *2*(2), 10-1128.

Gifford, E.W. (1919). Expedition of the California Academy of Sciences to the Galapagos Islands, 1905-1906. XIII. Field notes on the land birds of the Galapagos Islands and of Cocos Island, Costa Rica. *Proceeding California Academy of Sciences* Fourth Series, 2, 189-258.

Grant, P.R., Grant, B.R., Petren, K. and Keller, L.F. (2005). Extinction behind our backs: the possible fate of one of the Darwin's finch species on Isla Floreana, Galápagos. *Biological Conservation*, 122(3), 499-503.

Green, R.E., Cornell, S.J., Scharlemann, J.P. and Balmford, A. (2005). Farming and the fate of wild nature. *Science*, 307(5709), 550-555.

Grice, A.C. (2006). The impacts of invasive plant species on the biodiversity of Australian rangelands. *The Rangeland Journal*, 28(1), 27-35.

Guénette, J.S. and Villard, M.A. (2005). Thresholds in forest bird response to habitat alteration as quantitative targets for conservation. *Conservation Biology*, 19(4), 1168-1180.

Hawkes, C.V., Wren, I.F., Herman, D.J. and Firestone, M.K. (2005). Plant invasion alters nitrogen cycling by modifying the soil nitrifying community. *Ecology Letters*, 8(9), 976-985.

Heber, S., and Briskie, J.V. (2010). Population bottlenecks and increased hatching failure in endangered birds. *Conservation Biology*, 24(6), 1674-1678.

Hixon, M.A., Pacala, S.W., and Sandin, S.A. (2002). Population regulation: historical context and contemporary challenges of open vs. closed systems. *Ecology*, 83(6), 1490-1508.

Hong, P., Schmid, B., De Laender, F., Eisenhauer, N., Zhang, X., Chen, H., Craven, D., De Boeck, H.J., Hautier, Y., Petchey, O.L. and Reich, P.B. (2022). Biodiversity promotes ecosystem functioning despite environmental change. *Ecology Letters*, 25(2), 555-569.

Hood-Nowotny, R., Rabitsch, I., Cimadom, A., Suarez-Rubio, M., Watzinger, A., Yáñez, P.S., Schulze, C.H., Zechmeister-Boltenstern, S., Jäger, H. and Tebbich, S. (2023). Plant invasion causes alterations in Darwin's finch feeding patterns in Galápagos cloud forests. *Science of the Total Environment*, 895, 164990.

Ives, A.R. and Carpenter, S.R. (2007). Stability and diversity of ecosystems. *Science*, 317(5834), 58-62.

Jamieson, I.G. and Ryan, C.J. (2000). Increased egg infertility associated with translocating inbred Takahe (*Porphyrio hochstetteri*) to island refuges in New Zealand. *Biological Conservation*, 94, 107-114.

Jäger, H., Buchholz, S., Cimadom, A., Tebbich, S., Rodríguez, J., Barrera, D., Walentowitz, A., Breuer, M., Carrión, A., Sevilla, C. and Causton, C. (2015). Restoration of the blackberry-invaded *Scalesia* forest: Impacts on the vegetation, invertebrates, and birds. *Galapagos Report*, 2016, 145-149.

Jiguet, F., Devictor, V., Ottvall, R., Van Turnhout, C., Van der Jeugd, H. and Lindström, Å. (2010). Bird population trends are linearly affected by climate change along species thermal ranges. *Proceedings of the Royal Society B: Biological Sciences*, 277(1700), 3601-3608.

Keeler, K.H. and Joern, A. (1995). Population processes. In the Changing Prairie: North American Grasslands, Oxford Academic, 82-99.

Keller, L. and Waller, D. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17(5), 230-241.

Kirkpatrick, M. and Jarne, P. (2000). The effects of a bottleneck on inbreeding depression and the genetic load. *The American Naturalist*, 155(2), 154-167.

Koop, J.A., Causton, C.E., Bulgarella, M., Cooper, E. and Heimpel, G.E. (2021). Population structure of a nest parasite of Darwin's finches within its native and invasive ranges. *Conservation Genetics*, 22(1), 11-22.

Lafferty, K.D. and Kuris, A.M. (1999). How environmental stress affects the impacts of parasites. *Limnology and Oceanography*, 44(3part2), 925-931.

Lande, R. (1993). Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes. *The American Naturalist*, 142(6), 911-927.

Leuba, C., Tebbich, S., Nemeth, E., Anchundia, D., Heyer, E., Mosquera, D.A., Richner, H., Rojas Allieri, M.L., Sevilla, C., and Fessl, B. (2020). Effect of an introduced parasite in natural and anthropogenic habitats on the breeding success of the endemic little vermilion flycatcher *Pyrocephalus nanus* in the Galápagos. *Journal of Avian Biology*, 51(8).

Lima, M. and Jaksic, F.M. (1999). Survival, recruitment and immigration processes in four subpopulations of the leaf-eared mouse in semiarid Chile. *Oikos*, 343-355.

Litt, A.R., Cord, E.E., Fulbright, T.E. and Schuster, G.L. (2014). Effects of invasive plants on arthropods. *Conservation Biology*, 28(6), 1532-1549.

Lyons, K.G., Brigham, C.A., Traut, B.H. and Schwartz, M.W. (2005). Rare species and ecosystem functioning. *Conservation Biology*, 19(4), 1019-1024.

Loeschcke, A., Markert, A., Wilhelm, S., Wirtz, A., Rosenau, F., Jaeger, K.E. and Drepper, T. (2013). TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synthetic Biology*, 2, 22-33.

Loye, J. and Carroll, S. (1995). Birds, bugs and blood: avian parasitism and conservation. *Trends in Ecology & Evolution*, 10(6), 232-235.

Mosquera, D., Fessl, B., Anchundia, D., Heyer, E., Leuba, C., Nemeth, E., Rojas, M. L., Sevilla, C. and Tebbich, S. (2022). The invasive parasitic fly *Philornis downsi* is threatening Little Vermilion Flycatchers on the Galápagos Islands. *Avian Conservation Ecology*, 17(1), 6.

Paulay, G. (1994). Biodiversity on Oceanic Islands: Its Origin and Extinction1. *American Zoologist*, 34(1), 134-144.

Rentería, J.L., Gardener, M.R., Panetta, F.D., Atkinson, R., and Crawley, M.J. (2012). Possible impacts of the invasive plant *Rubus niveus* on the native vegetation of the *Scalesia* forest in the Galapagos Islands. *PLoS One*, 7(10), e48106.

Rentería, J., Atkinson, R., Crespo, C., Gardener, M. and Grosholz, E. (2021). Challenges for the management of the invasive blackberry (*Rubus niveus*) in the restoration of the *Scalesia* forest in the Galapagos Islands. *Invasive Plant Science Management*, 14(1), 20-28.

Ridgway, R. (1897). Birds of the Galapagos Archipelago. *Proceedings of the United States National Museum*, 19(1116), 459-670.

Robinson, S., Simmons, L. and Kennington, W. (2014) Sex-biased mortality associated with inbreeding in *Drosophila melanogaster*. *BMC Evolutionary Biology*, 14(1), 1-9.

Sekercioglu, C.H. (2002). Forest fragmentation hits insectivorous birds hard. *Directions in Science*, 1, 62-64.

Sigwart, J.D., Bennett, K.D., Edie, S.M., Mander, L., Okamura, B., Padian, K., Wheeler, Q., Winston, J.E. and Yeung, N.W. (2018). Measuring biodiversity and extinction—present and past. *Integrative and Comparative Biology*, 58(6), 1111-1117.

Snodgrass, R.E. and Heller, E. (1904). Papers from the Hopkins-Stanford Galapagos Expedition, 1898-1899. XVI Birds. *In Proceedings of the Washington Academy of Sciences*, 231-372.

Spigler, R.B., Theodorou, K. and Chang, S.M. (2017). Inbreeding depression and drift load in small populations at demographic disequilibrium. *Evolution*, 71(1), 81-94.

Stanton, R.L., Morrissey, C.A. and Clark, R.G. (2018). Analysis of trends and agricultural drivers of farmland bird declines in North America: A review. *Agriculture, Ecosystems & Environment*, 254, 244-254.

Steadman, D.W. (1995). Prehistoric extinctions of Pacific island birds: biodiversity meets zooarchaeology. *Science*, 267(5201), 1123-1131.

Swarth, H. (1931). The avifauna of the Galapagos Islands. *Occasional Papers of the California Academy of Sciences*, 18, 1-299.

Tallamy, D.W. and Shriver, W.G. (2021). Are declines in insects and insectivorous birds related? *Condor*, 123(1), duaa059.

Walentowitz, A., Manthey, M., Bentet Preciado, M. B., Chango, R., Sevilla, C., and Jäger, H. (2021). Limited natural regeneration of unique *Scalesia* forest following invasive plant removal in Galapagos. *PloS ONE*, 16(10), e0258467.

White, G.C. (2000). Modeling population dynamics. *Ecology and Management of Large mammals in North America*, 85-107.

Wolfe, J.D., Johnson, M.D. and Ralph, C.J., 2014. Do birds select habitat or food resources? Nearctic-Neotropic migrants in northeastern Costa Rica. *PloS One*, 9(1), e86221.

List of Publications

CHAPTER 1

Published

Anchundia, D. J., Green, R., Pike, C. L., Gutiérrez, G., Pibaque, P., Chango, R., Sevilla, C., Fessl, B., & Tebbich, S. (2024). Habitat restoration to conserve the Little Vermilion Flycatcher on Santa Cruz Island, Galapagos. *Bird Conservation International*, 34, 1-10. https://doi.org/10.1017/S0959270924000091

Author Contribution –**David Anchundia:** Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing – original Draft, Writing – review & editing, Visualization, Funding acquisition. **Rhys Green:** Methodology, Investigation, Formal analysis, Validation, Data Curation Writing – Review & Editing. **Courtney Pike:** Methodology, Formal analysis, Validation, Visualization, Writing – Review & Editing. **George Gutierrez:** Methodology, Data Curation, Writing – Review & Editing. **Peter Pibaque:** Methodology, Data Curation, Writing - Review & Editing. **Rafael Chango:** Methodology, Resources, Writing - Review & Editing. **Sevilla Christian:** Methodology, Resources, Writing - Review & Editing, Funding acquisition. **Birgit Fessl:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Funding acquisition. **Sabine Tebbich:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Funding acquisition.

CHAPTER 2

Published

Anchundia, D. J., Lam, A. W., Henderson, J. B., Van Dam, M. H., & Dumbacher, J. P. (2024). Genome assembly of *Pyrocephalus nanus*, a step toward the genetic conservation of the endangered Little Vermilion Flycatcher of the Galapagos Islands. *Genome Biology and Evolution*, 16(5), evae083. https://doi.org/10.1093/gbe/evae083

Author Contribution –**David Anchundia:** Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing – original Draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Athena Lam:** Methodology, Investigation, Resources, Writing – Review & Editing, Supervision. **James B. Henderson:** Methodology, Software, Validation, Formal

analysis, Data Curation, Writing – Review & Editing. **Matthew Van Dam:** Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Review & Editing, Visualization. **John Dumbacher:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing. Supervision, Project administration, Funding acquisition.

CHAPTER 3

Submitted

Anchundia, D. J., Lam, A. W., Henderson, J. B., Van Dam, M. H., Pike, C. L., Tebbich, S., Fessl, B., Richner, H., & Dumbacher, J. P. Submitted. Phylogeny, population structure, and conservation genomics of Galapagos vermilion flycatchers (genus *Pyrocephalus*). *Journal of Molecular Phylogenetic and Evolution*, Submitted 12 November 2024. http://dx.doi.org/10.2139/ssrn.5026862

Author Contribution –**David Anchundia:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – original Draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Athena Lam:** Methodology, Investigation, Resources, Writing – Review & Editing, Supervision. **James B. Henderson:** Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Review & Editing. **Matthew Van Dam:** Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Review & Editing, Visualization. **Courtney Pike:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Review & Editing, Visualization. **Sabine Tebbich:** Writing – Review & Editing, Supervision, Funding acquisition. **Birgit Fessl:** Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition. **Heinz Richner:** Conceptualization, Methodology, Writing – Review & Editing, Supervision. **John Dumbacher:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing. Supervision, Project administration, Funding acquisition.

CHAPTER 1

Habitat restoration to conserve the Little Vermilion Flycatcher *Pyrocephalus nanus* on Santa Cruz Island, Galapagos

DAVID J. ANCHUNDIA · RHYS GREEN · COURTNEY L. PIKE · GEORGE GUTIERREZ · RAFAEL CHANGO · CHRISTIAN SEVILLA · BIRGIT FESSL · SABINE TEBBICH

Photo by David Anchundia

Bird Conservation International (2024) 34: 1-10. https://doi.org/10.1017/S0959270924000091
Received: 04 August 2023 / Revised: 11 March 2024 / Accepted: 12 March 2024
Published online: 29 April 2024

Bird Conservation International

www.cambridge.org/bci

Habitat restoration to conserve the Little Vermilion Flycatcher *Pyrocephalus nanus* on Santa Cruz Island, Galapagos

David J. Anchundia[1,2], Rhys Green[3], Courtney L. Pike[1], George Gutiérrez[2], Peter Pibaque[2], Rafael Chango[4], Christian Sevilla[4], Birgit Fessl[2†] and Sabine Tebbich[1†]

[1]Department of Behavioral and Cognitive Biology, University of Vienna, Vienna, Austria; [2]Charles Darwin Research Station, Charles Darwin Foundation, Puerto Ayora, Santa Cruz, Galapagos, Ecuador; [3]Department of Zoology, University of Cambridge, Cambridge, UK and [4]Galapagos National Park Directorate, Puerto Ayora, Santa Cruz, Galapagos, Ecuador

Research Article

Cite this article: Anchundia DJ, Green R, Pike CL, Gutiérrez G, Pibaque P, Chango R, Sevilla C, Fessl B, Tebbich S (2024). Habitat restoration to conserve the Little Vermilion Flycatcher *Pyrocephalus nanus* on Santa Cruz Island, Galapagos. *Bird Conservation International*, **34**, e14, 1–10
https://doi.org/10.1017/S0959270924000091

Received: 04 August 2023
Revised: 11 March 2024
Accepted: 12 March 2024

Keywords:
Conservation *Pyrocephalus nanus*; Habitat management; Invasive species

Corresponding author:
David J. Anchundia;
Email: david.anchundia@fcdarwin.org.ec

†Joint last authors

Summary

The endemic Little Vermilion Flycatcher (LVF) *Pyrocephalus nanus* has suffered a drastic decline on Santa Cruz Island, Galapagos, where it was common 30 years ago. Currently, fewer than 40 individuals remain in the last remnants of natural humid forest in the Galapagos National Park on the island. This small population has low reproductive success, which is contributing to its decline in Santa Cruz. Previous studies have identified Avian Vampire Fly *Philornis downsi* parasitism, changes in food sources, and habitat alteration as threats to this species. In Santa Cruz, invasive plants may strongly affect the reproductive success of LVF because they limit accessibility to prey near the ground, the preferred foraging niche of these birds. Since 2019, we restored the vegetation in seven plots of 1 ha each by removing invasive blackberry plants and other introduced plant species. In all nests that reached late incubation, we also reduced the number of Avian Vampire Fly larvae. In this study, we compared foraging and perch height, pair formation, incubation time, and reproductive success between managed and unmanaged areas. As predicted, we found significantly lower foraging height and perch height in 2021 in managed areas compared with unmanaged areas. In 2020, the daily failure rate (DFR) of nests in the egg stage did not differ between management types; however, in 2021, the DFR in the egg stage was significantly lower in managed areas than in unmanaged areas. The DFR during the nestling stage was similar between managed and unmanaged areas in 2020, but in 2021, only nests in managed areas reached the nestling stage. Females brooded significantly more during the incubation phase in managed areas. Additionally, we found significantly higher reproductive success in managed areas compared with unmanaged areas in 2021, but not in 2020. Habitat restoration is a long-term process and these findings suggest that habitat management positively affects this small population in the long term.

Introduction

Recent studies raise concerns about the drastic decline of many avian species worldwide (BirdLife International 2022). One of the main groups of birds affected is the group that feeds on arthropods (Tallamy and Shriver 2021). Frequently identified causes for such declines include habitat loss and use of pesticides, which result in a reduction of arthropods, the birds' main food supply (Forister et al. 2019; Mahmood et al. 2016; Sekercioglu 2002; Wagner et al. 2021; Wauters et al. 2017). In the Galapagos Islands, declines of insectivorous birds are becoming more common (Dvorak et al. 2012, 2017, 2020). The original habitat of landbirds on several islands in the Galapagos has changed dramatically since colonisation by humans (Alomía et al. 2022; Geladi et al. 2022; Lundh 2006; Watson et al. 2010), altering food sources and adding further pressure on many vulnerable species.

The genus of birds with the most dramatic decline on the Galapagos Islands is *Pyrocephalus*, which originally included two endemic taxa, Little Vermilion Flycatcher *Pyrocephalus nanus* and Least Vermilion Flycatcher *Pyrocephalus dubius* (Carmi et al. 2016). The Least Vermilion Flycatcher, endemic to San Cristobal Island, is now considered "Extinct" (IUCN 2023); thus, it is the first extinction of an avian species in recent history in the Galapagos (Dvorak et al. 2020). Long-term monitoring of landbird populations has revealed that the Little Vermilion Flycatcher (hereafter LVF) is disappearing from several islands of the Galapagos archipelago. This species originally inhabited 10 large islands but has disappeared from two islands (Santa Fe and Floreana Islands) and is in sharp decline on Santiago Island (Fessl et al. 2017). The LVF is currently classified as "Vulnerable" in the IUCN Red List (IUCN 2023).

On Santa Cruz Island, the second largest island in the archipelago (986 km^2), LVF was abundant until 30 years ago (Merlen 2011). Since then, the population has decreased rapidly, with an estimate of fewer than 30–40 breeding pairs left in 2016 (Fessl et al. 2017). Small population sizes favour the mating of genetically close individuals, leading to inbreeding and increased hatching failure (Kruuk et al. 2002), which could add more stress to this population. Santa Cruz has the largest human population in the Galapagos and large areas of native forest have been cleared for agriculture (Alomía et al. 2022; Benitez-Capistros et al. 2014; Espinoza De Janon 2013, Geladi et al. 2022). The remaining LVF population in Santa Cruz has taken refuge in the forest in the humid zone (550–720 m a.s.l.) in a 7-km^2 area in the Galapagos National Park. Despite being protected, this area has been invaded by several invasive plant species such as blackberry *Rubus niveus*, guava *Psidium guajava*, orange cestrum *Cestrum auriculatum*, quinine *Cinchona pubescens*, cedrela *Cedrela odorata*, grasses *Paspalum conjugatum* and *Cenchrus purpureus*, and ferns *Dennstaedtia globulifera* and *Ctenitis sloanei*, reducing its habitat quality (Carrión-Klier et al. 2022; Rivas-Torres et al. 2018).

Recent studies comparing populations of LVFs on Santa Cruz and Isabela Island suggest that the decline in Santa Cruz is associated with altered arthropod diversity and accessibility due to habitat change (Leuba et al. 2020), caused by the invasive blackberry plant (Renteria et al. 2021). The blackberry is spreading aggressively within humid zones, where it competes with native and endemic plants, preventing them from regenerating (Renteria et al. 2012, 2021; Walentowitz et al. 2021). Blackberry brambles form a dense, tangled understorey up to 4 m in height (Renteria et al. 2012), leaving few open areas near the ground where LVFs can hunt for arthropods (Leuba et al. 2020). Due to restricted access to the ground on Santa Cruz, LVFs usually capture prey above invasive plants. However, in areas with no invasive blackberry on Isabela, they take their prey very close to the ground and have a higher foraging success than on Santa Cruz (Leuba et al. 2020).

Another major threat to LVFs is the invasive Avian Vampire Fly (Vampire Fly) *Philornis downsi* (Leuba et al. 2020; Mosquera et al. 2022). Its larval stages feed on the blood and tissue of nestlings, causing blood loss, anaemia, beak deformations, and nestling mortality (Fessl et al. 2018). When Vampire Fly parasitism was experimentally reduced in LVF nests, they had higher breeding success than in highly parasitised nests. This suggests Vampire Fly parasitism significantly decreases the LVFs' breeding success and may be one of the main factors influencing its population declines. Furthermore, the effects of Vampire Fly parasitism may be mediated by habitat quality as Leuba et al. (2020) found higher LVF breeding success and food provisioning rates in a more preserved site compared with a site with lower quality habitat.

However, Vampire Fly parasitism does not seem to be the main reason for breeding failure in Mina de Granillo Rojo in Santa Cruz. We found that 33 out of 56 nests were abandoned at an early stage of incubation, without any signs of predation or parasitism by the Vampire Fly (Leuba et al. 2020; Charles Darwin Foundation (CDF), unpublished). We hypothesise that the high percentage of nest abandonment and overall poor breeding success in Santa Cruz is linked to altered habitat quality (Kitaysky et al. 1999; Ouyang et al. 2012) caused by invasive plants, especially blackberry, which reduce prey accessibility.

To test this hypothesis, we carried out an experimental habitat management in seven plots of 1-ha each. In collaboration with the Galapagos National Park Directorate (GNPD), we cleared the invasive understorey from these plots and compared managed and unmanaged areas with respect to: (1) foraging height and perch height of LVF; (2) pair formation; (3) daily nest failure rate at egg and nestling stages; (4) incubation time; (5) breeding success. We predicted that managed areas would have lower foraging and perch height, more successful pair formation, lower nest failure rate, and higher breeding success than unmanaged areas.

Methods

Study area

This study took place in a cloud forest in the humid zone in the centre of Santa Cruz Island (90°21'30.69"W, 0°37'58.45"S) between 550 m and 720 m a.s.l. This area, called Mina de Granillo Rojo, covers approximately 700 ha, with 37 km of trails, which were used to search for birds and nests (Figure 1). The climate is seasonal, with a warm and humid climate from December to May and a cooler, dry climate during the remaining months.

Study species

The LVF has marked sexual dimorphism. The adult male has bright red and black feathers; however, during transition to adulthood, the immature male's plumage changes successively from yellow/brown (juvenile) to orange and a dull red colour (first-year breeding), with pronounced individual variation. In contrast, females have brown and yellowish plumage, with some individuals having sporadic orange feathers on the head. LVFs feed exclusively on arthropods (Guerrero and Tye 2011; Leuba et al. 2020), and prefer to forage by sallying from perches into open and semi-open areas both on the ground and in the air (Leuba et al. 2020). The birds build an open cup nest using lichens, moss, and tree bark as materials (D.J.A. personal observations) (Supplementary material Figure S1).

Habitat management

The plots for experimental management were selected using maps of vegetation types and previous sightings of LVFs. Seven plots of 1-ha each were selected for management. Plots were chosen based on accessibility and the presence of existing territories (Figure 1), not selected randomly. Invasive plants, including blackberry, orange cestrum, quinine, grasses, and ferns, were controlled and removed along with their roots to prevent regrowth. Workers used machetes and heavy cutting iron blades (*barretones*). Herbicides were not used to avoid any potential secondary impact on arthropods. From the beginning of the restoration in January 2019 to May 2021, each plot was cleaned every three to four months to prevent invasive plants from recolonising. The vegetation was cut at a height between 10 cm and 20 cm above the ground to mimic the vegetation structure at sites grazed by Galapagos Giant Tortoises *Chelonoidis niger*, which are the natural top herbivore in the Galapagos (Froyd et al. 2014; Hunter et al. 2021), but absent from our study area. This allowed native plants to germinate and grow faster.

Foraging and perch height

We collected data throughout the entire nesting season, which runs from December to May, when arthropod availability increases. To limit the effect of variation in foraging behaviour, observations took place at random, from 07h00 to 11h30, when the birds have the highest foraging activity. We did not conduct observations during

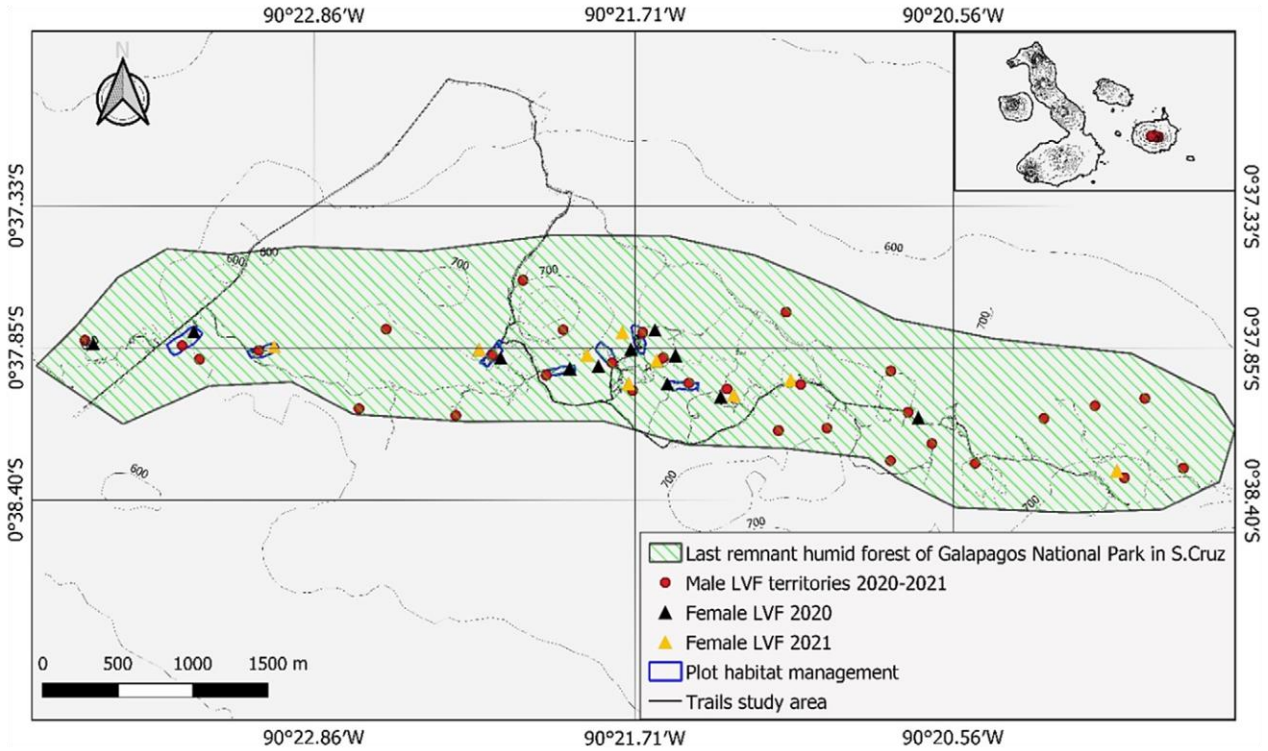


Figure 1. Study area and location of the last remnant of the Little Vermilion Flycatcher *Pyrocephalus nanus* population in Santa Cruz Island.

rainy weather because bird activity is usually limited. Upon entering an individual's territory, observers carefully followed the focal individual from afar using binoculars to avoid disturbance. The first time that an observed individual was seen foraging was referred to as a "first foraging observation". Each observation terminated after the first foraging attempt was recorded. Individuals were identified by plumage differences and by using combinations of colour rings; 11 males and six females were colour ringed during 2020 and 2021, which represents an estimated 50% of the males and females observed. Records of unringed breeding birds in a territory within one breeding season were assumed to be of the same individual. After a first foraging attempt (successful or not) during an observation, we recorded the height at which it foraged, the height of the perch from which the foraging sally was launched, and the type of surface where it captured its prey (e.g. ground, trunk, branches, leaves, air). The type of prey was recorded to the level of order if it was possible to identify it before the bird ate it.

Nest monitoring

Nest searching was conducted along 37 km of trails. These trails went through areas with managed and unmanaged habitat (Figure 1). In addition, narrow paths were opened in unmanaged areas to make these more accessible. Equal search effort was dedicated to managed and unmanaged areas. Five days per week, from 7h00 to 11h30, two teams of two people each simultaneously searched for birds and collected data in managed and unmanaged areas. The probability of detecting a territorial pair was high in both areas due to this intensive monitoring. Additionally, the behaviour of the LVF male is very noticeable, as they frequently fly above the canopy to display a territorial song. Also, males call the females during the mating phase, which was one of the main behaviours used to find an active territory. Pair formation was identified by the presence of an aggressive male and an aggressive female or a male and a female interacting and/or copulating. Breeding seasons spanned two successive calendar years. The season running from December 2019 to 15 March 2020 is referred to as "2020". The breeding season of the LVF continued until the end of May 2020, but it was not possible in 2020 to continue monitoring nests for the whole season due to the Covid-19 pandemic and GNPD restrictions. In the second year, "2021", data were collected during the whole breeding season from December 2020 to May 2021.

After an active nest was found, it was visited every three days during the incubation and feeding phases until its activity ended. Any effect that the observer could have on the behaviour of the adults near/at their nest was minimised as much as possible. During each visit, the nest was first observed with binoculars from a distance of 15–20 m and then filmed with an endoscopic camera when the birds left the nest to forage. The process of filming the open cup nests took ~20–30 seconds. By filming from above, we could count the number of eggs or nestlings and determine whether there were any changes since the previous visit. After filming, we waited at a distance until the female returned to the nest. Our methodology followed Leuba et al. (2020), which found no abandonment attributed to observations and nest filming. The total number of active days for a nest was determined by the start of the incubation phase to the last date of activity.

Incubated nests (42 nests: 25 in 2020, 17 in 2021) were observed every three days for one hour to record the total time the female was inside the nest. We could not compare parental feeding rate of nestlings between managed and unmanaged territories because most nests in unmanaged areas were abandoned during incubation. The outcome of each nest was categorised as follows. (1) Fledged (at least one nestling fledged) or failed; a fledgling was defined as a

nestling that left the nest and stayed around the nest waiting to be fed by the parents. Failed nests during the incubation phase were categorised as (2) Abandoned: when eggs were intact in the nest, (3) Empty during the incubation phase; a nest previously containing eggs and found empty (ejection by wind/rain), or (4) Predated: a destroyed or partially destroyed nest, and/or broken eggshell remains. As nests of the LVF are fragile and small (8–9 cm diameter) in comparison to all potential predators, predation events leave visible destruction of the nest (Leuba et al. 2020; D.J.A. personal observations).

Failed nests during the nestling phase were categorised as follows. (5) Empty during nestling phase: a nest previously containing nestling(s) and found empty before the end of the nestling period (ejection by wind or removal of dead nestlings by parents can be causes of this). (6) Dead nestling(s): one or more nestlings found dead in the nest. (7) Predated: remains of nestlings' body parts or remains of feathers or down (see above), and/or destroyed or partially destroyed nest. (8) Unknown: nests that failed due to unknown reasons. For five nests, we could not establish the reason of failure, as we could not enter the field site during the weeks they failed due to pandemic regulations. Once we confirmed on three nest visits that activity had ceased, each nest was collected for a detailed examination.

Eggs

If a nest failed or was abandoned, the eggs were collected to check whether they were fertile and to see if abandonment was likely due to egg infertility. We candled eggs using a flashlight in a darkened room to check for the presence of embryos. However, if an egg contained a small embryo, this method was unreliable, so we opened eggs with no visibly large embryo and inspected the contents carefully in a Petri dish.

Control of Vampire Fly infestation in LVF nests

To minimise the influence of Vampire Fly parasitism on breeding success, we reduced the number of larvae from all nests that reached late incubation by injecting Permacap® (Controlled Release Permethrin) into the nest. Permacap is an insecticide used to target and kill Vampire Fly larvae in nests of Galapagos landbirds (Cimadom et al. 2019; Leuba et al. 2020; Mosquera et al. 2022). Approximately 5 ml of a 0.5% (by volume) aqueous solution of Permacap was injected into the base of the nest. A second injection of 1–3 ml of Permacap was applied if some of the solution ran off from the nest during the initial application. The solution was applied using a syringe secured to the end of a carbon fibre tube extending up to 7 m. The solution in the syringe was pushed out by a pressurised air pump. Each nest was treated in the late stage of incubation from days 11 to 15 (mean number of days of incubation for LVF is 16 days). Injection of the insecticide solution was carried out while standing on a ladder for closer proximity to each nest. No nest abandonment was observed after the treatment.

Of the 42 nests monitored, only 20 nests (15 nests in managed and five nests in unmanaged areas) were treated against Vampire Fly parasitism because the other nests did not reach the late incubation stage. After nest activity ceased, nests were collected, dismantled in the laboratory, and the number of Vampire Fly pupae or larvae and their larval stage were recorded. Treated nests had an average of 1.80 ± SD 3.8 larvae/nest. In previous studies on this species in an unmanaged, anthropogenic habitat (El Cura, Isabela), the average number of larvae was 16.4 ± SD 9.7 (Leuba et al. 2020).

Of the treated nests, 15 nests had no detectable Vampire Fly parasites when examined after the end of the breeding attempt and 5 nests had Vampire Fly parasites (3, 4, 5, 10, 14 larvae or pupae). Of those five nests, only one was identified as possibly failing due to Vampire Fly parasitism (with 14 larvae). Two nests that had 10 and three Vampire Fly larvae were successful, while two nests with low numbers of Vampire Fly larvae (four and five) were probably destroyed by wind and heavy rain. Of the 22 nests that were not treated because they did not reach the late incubation stage (15 managed, seven unmanaged), 17 nests did not have Vampire Fly larvae. The other five nests that reached mid incubation stage had Vampire Fly larvae (2, 2, 4, 7, and 10 larvae); all five nests were in managed areas. The one nest with 10 larvae failed during heavy rain conditions and it is not known with certainty whether parasitism or rain influenced the abandonment.

Statistical analysis

Foraging and perch height

We ran robust linear mixed models (RLMMs) using the "rlmer" function from the *robustlmm* package (Koller 2016) in R, version 4.2.1 (R Core Team 2022) to analyse foraging height and perch height as dependent variables in separate models. For the analysis, "first foraging observation" data were used and foraging observations with incomplete data were excluded. For fixed effects, management (binary variable: yes/no) was included as the main independent variable and sex was included as a binary covariate. Individual ID was modelled as a random factor "to account for pseudoreplication". This model was chosen due to non-normality of residuals and heteroscedasticity. Only data from 2021 were used in this analysis because these data were collected during the entire breeding season.

Pair formation

We used a Fisher's exact test for a 2 × 2 table to compare the number of paired and unpaired males in managed and unmanaged areas. Only data from 2021 were used in this analysis because these data were collected during the entire breeding season.

Daily failure rate at egg and nestling stages

We examined the effects of habitat management on the daily probability of failure of nests, the daily failure rate (DFR), at the egg and nestling stages using an extension of the method of Mayfield (1975), which allows for variation among nests and nesting stages in the period during which the outcome was monitored. We followed Aebischer (1999) in using a generalised linear model (GLM) with a logistic link function and binomial error term to fit logistic regression models in which the number of binomial trials in the analysis is the number of days for which each nest was monitored and on which its outcome (survival or failure) on each day was observed. The total number of binomial trials per nest was therefore the number of exposure-days of the nest in the Mayfield method; for further details see Aebischer (1999). We adapted the approach by taking the binary dependent variable to be the outcome on each exposure-day to be failure (1 = failed) or survival (0 = survived). Exposure-days and failure events for each nest were partitioned between the egg stage and nestling stage and a binary independent variable representing stage (0 = egg stage; 1 = nestling stage) was included in the model for each exposure-day to account for the possible effect of stage on DFR. For the purposes of modelling, we assumed that nests that failed did so at the midpoint between the penultimate and final nest check. Our principal

objective was to model DFR in relation to two independent variables, i.e. habitat management and breeding season, both recorded as binary variables (unmanaged = 0, managed = 1, 2020 = 0, 2021 = 1).

Studies of DFR often find differences between the egg stage and the nestling stage, so we would ideally have fitted separate models to data from these two stages. However, this was not possible because no nesting attempts survived to hatching in unmanaged territories in 2021. Instead we analysed egg-stage and nestling-stage data together in the same models using the logistic regression method described above and assumed that the effect of stage on logit (DFR) was independent of season and management treatment. We fitted five models including main effects, in addition to stage, of (1) season or (2) habitat management included separately, (3) both of these main effects included together, (4) both of the main effects and their two-way interaction, and (0) the null model with no covariates except for stage. For each fitted model, we used the residual deviance and number of fitted parameters to calculate the small-sample Akaike Information Criterion (AIC_c) (Burnham and Anderson 2002). We selected the model with the lowest AIC_c value.

The survival of breeding attempts from the first-egg date (FED) to hatching and from hatching to fledging was calculated by raising the daily attempt survival rates (1-DFR) estimated from Model 4 to the power of the mean duration in days of FED to hatching (16.8 days) and hatching to fledging (20.4 days) periods. These two survival rates were multiplied together to give the probability of survival of a nesting attempt from FED to fledging. We multiplied these probabilities by the mean number of young per fledged brood to obtain the mean number of fledglings per breeding attempt. There were insufficient observations of the size of fledged broods to calculate means separately for each year and management treatment category, so we multiplied by the mean number of young per fledged brood for both treatments combined, which was 2.0 in both seasons.

Incubation time per hour

We fitted a linear least-squares mixed model (LMER) in R using the *lme4* package (Bates et al. 2015) with female attendance in nest (in minutes per hour) as the dependent variable and management type and year as independent variables modelled as fixed effects, both being taken to be binary variables (untreated = 0, treated = 1, 2020 = 0, 2021 = 1). Nest ID was modelled as a nested random factor with pair ID to account for pseudoreplication. A variance inflation factor (VIF) test was conducted to test for collinearity among predictors. Additionally, the model was checked for normally distributed residuals using the Shapiro–Wilk test for normality and by visually observing the Q–Q plot of the residuals. Homoscedasticity of residuals was also confirmed via visual inspection of scaled residuals and by conducting a studentised Breusch–Pagan test using the "bptest" function from the *lmtest* package (Zeileis and Hothorn 2002). *T* values and *P* values were obtained using the *lmerTest* package (Kuznetsova et al. 2017).

Season-long breeding productivity

The preceding analyses allowed us to estimate the mean number of fledglings per breeding attempt, but individual pairs were observed to make several breeding attempts in the same season, often replacing failed attempts. Ideally, we would also have estimated season-long breeding productivity by using a simulation model of nesting (Beintema and Muskens 1987; Green et al. 1997), but there were insufficient data to estimate all the relevant parameters separately for the two seasons and two habitat treatments. We therefore divided the total number of fledglings recorded for each season per treatment combination by the number of territories monitored to give the mean season-long productivity. We expect this approach to be accurate, even if we did not find and record every breeding attempt in every territory. That is because breeding attempts that survived the 37 days from FED until fledging would be highly likely to be detected at some time during this lengthy period.

Of the nesting attempts that were recorded as hatching ($n = 12$), the first record of eggs being present in the nest was obtained within four days of FED in all cases (mean 1.7 days; range 0–4 days), even though the duration of incubation was 17 days. Although sample sizes were small, there was no indication of any marked difference in this interval between nests in unmanaged and managed plots (unmanaged: mean 2.5 days; range 1–4 days; $n = 2$: managed: mean 1.5 days; range 0–3 days; $n = 10$). In addition, nests were often found during building before eggs were laid. We concluded that our simple measure of season-long productivity is likely to have been accurate for both unmanaged and managed territories.

Results

Foraging and perch height

Of 242 foraging observations of males and females, 160 observations (of 10 individuals) were in managed areas and 82 observations (of 15 individuals) were in unmanaged areas. We recorded the foraging height of 237 observations and perch height of 236 observations. The RLMM showed that management type had a significant effect on foraging height ($t = -3.72$, $P < 0.001$, $n = 237$), with birds having a 1.39 ± SE 0.37 m lower foraging height in managed areas. Perch height was also significantly lower in managed areas ($t = -3.87$, $P < 0.001$, $n = 236$), with a 1.16 ± SE 0.30 m lower perch height in managed areas (Figure 2). Sex of individuals did not significantly affect foraging height ($t = 0.47$, $P = 0.64$), nor perch height ($t = 0.69$, $P = 0.49$). Of the 242 foraging observations, we were able to identify 70 prey items as Lepidoptera in larval stage (caterpillars) (36% managed, 16% unmanaged), followed by 31 prey items as adult stage Lepidoptera (moths/butterflies) (12% managed, 13% unmanaged), two prey items as Diplopoda (millipedes), and one item as Coleoptera (ladybugs) (2% managed, 1% unmanaged). A total of 138 observations that were categorised as unknown prey (50% managed, 71% unmanaged) were small invertebrates that were immediately swallowed by the bird when hunting in the air.

Pair formation

In managed areas, we found four males with a mate and three males without a mate, while in unmanaged areas, we found five males with a mate and 11 males without a mate. Two of these pairs in unmanaged areas did not nest. There was no significant difference in pair formation between managed and unmanaged areas (Fisher's exact test, two-tailed, $P = 0.36$).

DFR at the egg stage

DFR at the egg stage was similar in the two management types in 2020, but was much higher in managed than in unmanaged territories in 2021 (Table 1). Nestling-stage DFR was very similar in unmanaged and managed territories in 2020. No nests were

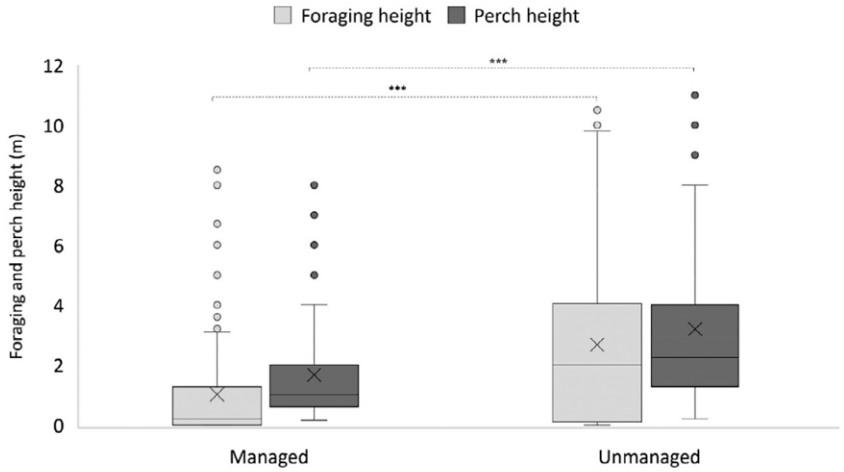


Figure 2. Foraging height and perch height of Little Vermilion Flycatcher *Pyrocephalus nanus* in managed versus unmanaged areas. The coloured circles in the graph represent the outliers of the collected data. The upper boundary of the boxplot reflects the third quartile, the lower boundary reflects the value of the first quartile, and the whisker lines show the least and greatest values. The line inside the boxplot reflects the median and the x symbol denotes the mean.

Table 1. Breeding success of Little Vermilion Flycatcher *Pyrocephalus nanus* on Santa Cruz in 2020 and 2021 in relation to habitat management in territories to control invasive plant species. DFR values are from stage-specific data for each treatment class and year and not from logistic regression models. No eggs hatched in monitored nets in 2021 in unmanaged territories, so the DFR could not be estimated. The probability of a breeding attempt surviving from FED to fledging was calculated from Model 4 (see text). The number of fledglings recorded per monitored territory recorded during the whole season is also shown. DFR = daily failure rate; FED = first-egg date

Parameter	2020		2021	
Habitat management	No	Yes	No	Yes
Breeding territories monitored	3	7	3	4
Attempts monitored at egg stage	7	18	5	12
Egg–stage DFR	0.0806	0.0733	0.1852	0.0337
Nestling–stage DFR	0.0500	0.0476	–	0.0263
Attempt survival probability: FED to fledge	0.0556	0.0745	0.0000	0.3033
Mean number of fledglings per attempt	0.111	0.149	0.000	0.607
Total fledglings recorded per territory	1.000	0.429	0.000	2.000

monitored at the nestling stage in unmanaged territories in 2021, so unmanaged and managed territories could not be compared. Logistic regression modelling identified the model of DFR at both nesting stages with main effects of breeding season, habitat management, and their two-way interaction (Model 4) as having the lowest AIC_c value of the five models compared (Table S1). The 95% confidence interval (CI) for the interaction term between season and management treatment provided strong evidence for a difference in the effect of treatment between seasons (95% CI for the interaction term for logit DFR: -3.318 to -0.180), there being no evidence for an effect of treatment on logit DFR in 2020 (95% CI for the treatment term: -1.072 to 0.881), but strong evidence for a negative effect of treatment in 2021 (95% CI for the treatment term: -3.135 to -0.613).

DFR at the nestling stage

Nestling-stage DFR was very similar in unmanaged and managed territories in 2020. No nests were monitored at the nestling stage in unmanaged territories in 2021, but nestling-stage DFR was considerably lower in managed territories in 2021 than in 2020. However, the sample sizes were small and this apparent difference between years was non-significant ($\chi^2_{(1)} = 0.36$, $P = 0.55$) (Table 1). Overall, nests in managed areas were active longer (mean 15.77 ± SE 2.00 days, $n = 30$) than nests in unmanaged areas (mean of 8.25 ± SE 2.94 days, $n = 12$ (Figure 3).

Incubation time

In areas with vegetation management, females spent significantly more time in the nest compared with unmanaged areas ($t = 2.74$, $P = 0.018$, $n = 116$), namely on average 12.18 ± SE 4.45 minutes longer per hour of observation (Figure 4). Year did not have a significant effect on the time the female spent in the nest ($t = -0.88$, $P = 0.40$, $n = 116$) (Figure 3).

Per-attempt nest success and season-long breeding productivity

As a consequence of the results obtained from the DFR analysis, per-attempt nest success was similar in unmanaged and managed territories in 2020, but there was a large difference in 2021, with

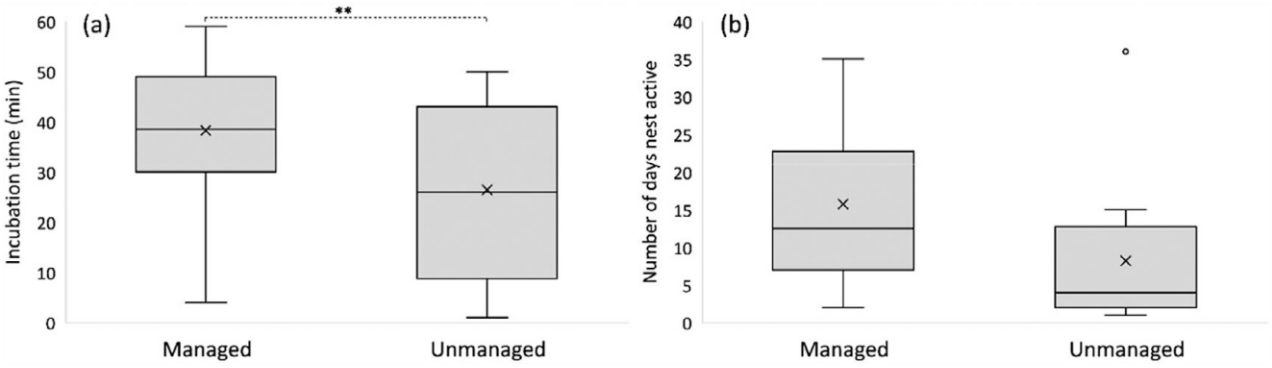


Figure 3. (a) Length of time (in minutes per hour) the female spent sitting on the nest for each management type. (b) Number of days that the nests were active in habitat managed versus unmanaged areas during the 2020 and 2021 breeding seasons combined. The upper boundary of the boxplot reflects the third quartile, the lower boundary is the value of the first quartile, the whisker lines indicate the least and greatest values. The line inside the boxplot reflects the median and the x symbol denotes the mean.

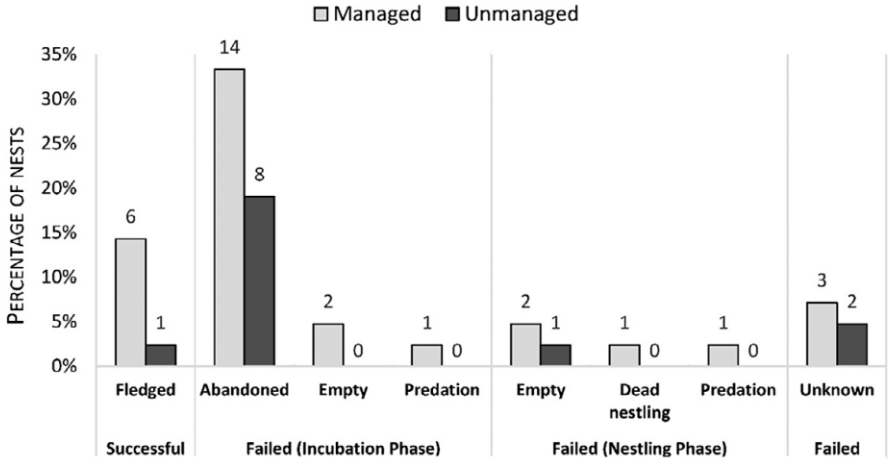


Figure 4. Nest outcome of Little Vermilion Flycatcher *Pyrocephalus nanus* in Santa Cruz during the 2020 and 2021 breeding seasons.

success being zero in unmanaged territories and substantially higher in managed territories than it was in managed territories in 2020 (Table 1). A similar pattern was seen for season-long breeding productivity (Table 1).

Nest outcome

In total, we observed 42 breeding attempts (30 in managed and 12 in unmanaged areas) from at least 10 breeding pairs. Not all the birds in our study were ringed, therefore, it could not be verified if they were the same breeding pairs in both years. Overall, seven successful nests were monitored during both years with 14 fledglings in total, including six nests in managed areas and one nest in an unmanaged area (Figure 4). Abandonment (with intact eggs) during the incubation phase was the main reason for breeding failure in both managed and unmanaged areas (Figure 4). During the nestling phase, one nest was found with a dead nestling and Vampire Fly parasitism and/or heavy rain probably caused this nest to fail. Predation seemed to play only a minor role, as only two nests showed obvious signs of predation (one nest in the incubation phase and one nest in the nestling phase). However, a more detailed study with trap cameras would be important in order to assess the level of predation.

Eggs

Most (60%) of the 78 eggs that were laid were fertile in both breeding seasons. Only one egg during those seasons was infertile. The fertility of 30 unhatched eggs could not be determined. Of these, 18 eggs were lost when they fell from the nests during nest collection from high tree branches and could not be found on the forest floor and 12 eggs were in early development, so we could not determine their fertility.

Discussion

Habitat management had a delayed positive effect on the breeding success of LVF. Although habitat management did not have a significant effect on pair formation, pairs that were in managed areas nests survived longer, pairs had more breeding attempts, longer incubation bouts, and higher breeding success in 2021. We hypothesise that this was mainly influenced by increased food availability in the managed areas, since we recorded a significantly lower perch and foraging height in managed territories in 2021. These results are similar to those of Leuba et al. (2020), who reported that a lower perch height in the near pristine site Alcedo was associated with higher foraging success. It would have been

valuable in our study to also measure foraging success as prey captures per unit time or the provisioning rate to nestlings, but we were unable to collect sufficient good quality data to do this.

Forests that have open areas, whether they are man-made or natural, generally have higher numbers of flycatchers than closed forest areas (Mannan 1984), since flycatchers generally prefer open areas to closed forests with dense vegetation (Beedy 1981; Mannan 1984). This may be due to the flycatchers' need for areas with perches and unobstructed airspace, which allows them to visually scan a large area for invertebrates on the ground or in the understorey (Beedy 1981). Given this, in open areas, researchers may be able to detect LVFs better and their conspicuous display behaviour, their tendency to approach humans, and their bright coloration increase detectability. Although the dense bramble thickets in unmanaged areas may have reduced our ability to detect unpaired females, our intensive search effort and the conspicuous behaviour of paired males render it unlikely that we missed breeding attempts or that our estimate of productivity was affected by the different visibility in managed and unmanaged areas. The detectability of birds during foraging is unlikely to have affected foraging height data. We understand that further concerns could arise with data from unmanaged areas where dense vegetation may have prevented sightings of foraging birds. However, we made detailed observations on LVF behaviour in areas with dense vegetation in prior years (Charles Darwin Foundation, unpublished data), which showed that LVFs preferred to move quickly to hunt by surprise, flew quickly or perched over vegetation, and foraged above it. They were not observed trying to manoeuvre their way through the dense vegetation.

Food availability plays an important role in bird breeding decisions and behaviour (O'Brien and Dawson 2011; Rubenstein 2007). In line with these studies, we recorded significantly more breeding attempts, longer nest duration, and longer incubation bouts in managed areas. In House Sparrows *Passer domesticus*, females prolonged their bouts of attendance at their nests in habitats with more food sources compared with areas with fewer food sources (Václav et al. 2003). Higher food density in territories may allow adults to avoid travelling far in the search for food and reduce the chance of extra pair copulation, which may also influence the time invested in the nest (Davies and Lundberg 1984). Likewise, a higher density of food may allow adults to increase their food delivery to nestlings and increase nest success.

Habitat restoration and effect on breeding success

Even though we found a positive effect of habitat management on DFR at the egg stage and on the probability that a nest would survive from egg laying to fledging, this effect was not apparent until 2021. This might be because regeneration of Galapagos native and endemic plants and their associated invertebrate fauna takes time after invasive plant removal (Cimadom et al. 2019). Management of habitat is a long-term investment and several forest restoration projects reported that birds increase their reproduction in these areas over time (Gaines et al. 2007; Hartung and Brawn 2005; Ortega-Álvarez and Cisneros 2012).

Despite habitat restoration efforts, many LVF nests failed during the early incubation phase in managed areas in the 2020 season, continuing a similar pattern observed since 2017. The first weeding of plots started in January 2019 and, during that whole year, there was more disturbance at the study site. Therefore, we initiated this study in 2020 to allow the native vegetation to regrow. However, it is likely that in 2020, nest abandonment was higher in managed areas as the vegetation and associated invertebrates were likely still recovering from the initial cutting of vegetation. In 2021, we did not record nest abandonment in early incubation and the birds remained very active and committed to their nests. This could suggest an improvement in habitat conditions over time in managed areas.

Weather and predation also played a role in nest failure. In both breeding seasons, torrential rains and strong winds affected the success of six LVF nests, including five nests in managed areas and one in an unmanaged area. Strong wind and rain can have a negative effect on passerine nests in the Galapagos, as the nests become wet and the eggs and nestlings get cold quickly (Cimadom et al. 2014). Additionally, Heyer et al. (2021) found that heavy rains can prevent foraging and nestling provisioning in Darwin's finches. Also, during the two years of monitoring, only two out of 42 nests were predated, suggesting predation only has a minor effect, as reported in prior studies (e.g. Leuba et al. 2020).

Despite factors influencing nest failure in both managed and unmanaged areas, currently the highland Scalesia forest may be the most suitable habitat for LVFs. In the past, LVFs were distributed throughout Santa Cruz Island in different and more open types of ecosystems (Merlen 2011; Rothschild and Hartert 1899). However, habitat reduction, highly intensive agriculture, and use of pesticides might have reduced prey sources. Additionally, another possible reason why the Santa Cruz LVF population persists in the highlands is that there is a lower Vampire Fly parasitism intensity at the beginning of the breeding season in higher, colder areas (Mosquera et al. 2022), which could have alleviated the pressures of parasitism.

Fertility and small population size

Contrary to expectations that the fertility of a small population would decrease due to a bottleneck effect, such as inbreeding (Jamieson and Ryan 2000), this appears not yet to be a frequent problem in this population since most eggs were fertile. The small number of individuals (30 birds in 2020 and 32 in 2021), with twice as many males as females observed, is of great concern. Perhaps some females or males were not detected, which could have affected the sex ratio. However, half of the males observed in our study were unmated and competed to mate with the few females we observed, which suggests that there are a limited number of females at the study site. Small populations are at higher risk of experiencing random imbalance of their sex ratio due to demographic and environmental stochasticity (Frankham and Wilcken 2006; Robinson et al. 2014). Male LVFs' distinct red plumage and courtship/territorial behaviours can make them more vulnerable to natural predators; however, this could also increase detectability for researchers. Contrastingly, females have cryptic plumage and are more evasive; however, they are more vulnerable while brooding in open cup nests, where predators could attack more easily. In general, for many threatened avian species, females have higher mortality rates than males (reviewed in Donald 2007). Females spend a lot of energy, laying up to five clutches with two to three eggs per season. Most of these nests were not successful, but they continued to make repeated attempts. This high-energy investment by the female required for egg development (Meijer et al. 1989) and the reduction of favourable food sources could influence their long-term survival and overall fitness (Hanssen et al. 2005).

Given the positive effect of habitat restoration in this study, habitat management and removal of invasive plants appears to be a valuable tool for the recovery of the LVF population. Based on our study and knowledge of the pressures on this population, we believe

that the most important management objectives for improving the breeding success and stabilising the population are as follows: (1) continued persistent control of Vampire Fly parasitism; (2) restoration of endemic forests without the use of pesticides and avoidance of additional habitat reduction; (3) continued controlling of any potential predation from invasive species, such as rats, cats, and possibly the invasive Smooth-billed Ani *Crotophaga ani*; (4) evaluation of the impact of diseases on LVF; (5) reduction in mortality from car impacts on birds; (6) evaluation of the consequences of the bottleneck effect on this population; (7) after the factors that are causing the LVF's decline are controlled, the evaluation and review of the possibility of translocation of individuals from other populations to Santa Cruz to increase genetic diversity if proven necessary.

Acknowledgements. This study was conducted with permission of the GNPD research permits PC-30-19, PC-11-20, and PC-06-21. We thank CDF plot maintenance workers Novarino Castillo and Klever Abad, CDF volunteers Keiko Cruz, Ruben Gil, and Peter VanMunnen; CDF scientists Carolina Carrión, Heinke Jäger, and Eileen Heyer; and several GNPD rangers. Thanks to John Dumbacher and Stefan Fischer for sharing their comments to improve the manuscript. This publication is contribution number 2545 of the Charles Darwin Foundation for the Galapagos Islands. The study was funded by a PhD fellowship awarded to David Anchundia from the World Wildlife Fund, Mr Russell E. Train and Education for Nature, Agreement #RH94, and a grant awarded to Sabine Tebbich and Rebecca Hood-Nowotny by the Austrian Science Fund (grant number P 32555-B). Other funding sources included the Galapagos Invasive Species Fund, Galapagos Conservation Trust, Galapagos Conservancy, Lindblad Expeditions – National Geographic, and Peter and Kris Norvig.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0959270924000091.

References

Aebischer N.J. (1999). Multi-way comparisons and generalized linear models of nest success: extensions of the Mayfield method. *Bird Study* **46**, S22–S31.

Alomía Herrera I., Paque R., Maertens M. and Vanacker V. (2022). History of land cover change on Santa Cruz Island, Galapagos. *Land* **11** 1017.

Bates D., Maechler M., Bolker B. and Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.

Beedy E.C. (1981). Bird communities and forest structure in the Sierra Nevada of California. *The Condor* **83** 97–105.

Beintema A.J. and Muskens G.J.D.M. (1987). Nesting success of birds breeding in Dutch agricultural grasslands. *Journal of Applied Ecology* **24**, 743–758.

Benitez-Capistros F., Hugé J. and Koedam N. (2014). Environmental impacts on the Galapagos Islands: Identification of interactions, perceptions and steps ahead. *Ecological Indicators* **38**. 113–123.

BirdLife International (2022). *State of the World's Birds 2022: Insights and Solutions for the Biodiversity Crisis*. Cambridge: BirdLife International.

Burnham K.P. and Anderson D.R. (2002). *Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach*. New York: Springer.

Carmi O., Witt C.C., Jaramillo A. and Dumbacher J.P. (2016). Phylogeography of the Vermilion Flycatcher species complex: Multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. *Molecular Phylogenetics and Evolution* **102**, 152–173.

Carrión-Klier C., Moity N., Sevilla C., Rueda D. and Jäger H. (2022). The importance of very-high-resolution imagery to map invasive plant species: Evidence from Galapagos. *Land* **11**, 2026.

Cimadom A., Jäger H., Schulze C.H., Hood-Nowotny R., Wappl C. and Tebbich S. (2019). Weed management increases the detrimental effect of an invasive parasite on arboreal Darwin's finches. *Biological Conservation* **233**, 93–101.

Cimadom,A., Ulloa A., Meidl P., Zöttl M., Zöttl E., Fessl B. et al. (2014). Invasive parasites, habitat change, and heavy rainfall reduce breeding success in Darwin's finches. *PLOS ONE* **9**, e107518.

Davies N.B. and Lundberg A. (1984). Food distribution and a variable mating system in the dunnock, *Prunella modularis*. *Journal of Animal Ecology* **53**, 895–912.

Donald P.F. (2007). Adult sex ratios in wild bird populations. *Ibis* **149**, 671–692.

Dvorak M., Fessl B., Nemeth E., Anchundia D., Cotin J., Schulze C. H. et al. (2020). Survival and extinction of breeding landbirds on San Cristóbal, a highly degraded island in the Galápagos. *Bird Conservation International* **30**, 381–395.

Dvorak M., Fessl B., Nemeth E., Kleindorfer S. and Tebbich S. (2012). Distribution and abundance of Darwin's finches and other land birds on Santa Cruz Island, Galapagos: evidence for declining populations. *Oryx* **46** 78–86.

Dvorak M., Nemeth E., Wendelin B., Herrera P., Mosquera D., Anchundia D. et al. (2017). Conservation status of landbirds on Floreana: the smallest inhabited Galapagos Island. *Journal of Field Ornithology* **88**, 132–145.

Espinoza De Janon, F. (2013). *Evaluación del Cambio de Uso de Suelo y Presencia de Especies de Flora Invasora en la Perimetral Agrícola de la isla Santa Cruz*. Bachelor's thesis, Universidad de Especialidades Espíritu Santo, Guayaquil.

Fessl B., Anchundia D., Carrion J., Cimadom A., Cotín J., Cunninghame F. et al. (2017). Galapagos landbirds (passerines, cuckoos, and doves): Status, threats and knowledge gaps. In *Galapagos Reports 2015–2016*. Puerto Ayora, Galapagos: GNPS, GCREG, CDF and GC, pp. 149–160.

Fessl B., Heimpel G.E. and Causton C.E. (2018). Invasion of an avian nest parasite, *Philornis downsi*, to the Galapagos Islands: colonization history, adaptations to novel ecosystems, and conservation challenges. In Parker P. (ed.), *Disease Ecology: Galapagos Birds and Their Parasites*. Cham: Springer, pp. 213–266.

Forister M.L., Pelton E.M. and Black S.H. (2019). Declines in insect abundance and diversity: We know enough to act now. *Conservation Science and Practice* **1**, e80.

Frankham R. and Wilcken J. (2006). Does inbreeding distort sex-ratios? *Conservation Genetics* **7**, 879–893.

Froyd C.A., Coffey E.E., van der Knaap W.O., van Leeuwen J.F., Tye A. and Willis K.J. (2014). The ecological consequences of megafaunal loss: giant tortoises and wetland biodiversity. *Ecology Letters* **17**, 144–154.

Gaines W.L., Haggard M., Lehmkuhl J.F., Lyons A.L. and Harrod R.J. (2007). Short-term response of land birds to ponderosa pine restoration. *Restoration Ecology* **15**, 670–678.

Geladi I., Henry P.Y., Couenberg P., Welsh R. and Fessl B. (2022). What are farmers' perceptions about farmland landbirds? A Galapagos Islands perspective. *Renewable Agriculture and Food Systems* **37**, 504–515.

Green R.E., Tyler G.A., Stowe T.J. and Newton A.V. (1997). A simulation model of the effect of mowing of agricultural grassland on the breeding success of the corncrake (*Crex crex*). *Journal of Zoology* **243**, 81–115.

Guerrero A.M. and Tye A. (2011). Native and introduced birds of Galapagos as dispersers of native and introduced plants. *Ornitología Neotropical* **22**, 207–217.

Hanssen S.A., Hasselquist D., Folstad I. and Erikstad K.E. (2005). Cost of reproduction in a long-lived bird: incubation effort reduces immune function and future reproduction. *Proceedings of the Royal Society London B: Biological Sciences* **272**, 1039–1046.

Hartung S. and Brawn J. (2005). Effects of savanna restoration on the foraging ecology of insectivorous songbirds. *The Condor* **107**, 879–888.

Heyer E., Cimadom A., Wappl C. and Tebbich S. (2021). Parental care in the small tree finch *Camarhynchus parvulus* in relation to parasitism and environmental factors. *Ibis* **163**, 137–149.

Hunter E., Blake S., Cayot L. and Gibbs J. (2021). Role in ecosystems. In Gibbs J. P., Cayot L.J. and Tapia Aguilera W. (eds), *Galapagos Giant Tortoises*. Cambridge: Academic Press/Elsevier, pp. 299–315.

IUCN (2023). *The IUCN Red List of Threatened Species. Version 2022-2*. Gland: International Union for Conservation of Nature.

Jamieson I. and Ryan C. (2000). Increased egg infertility associated with translocating inbred takahe (*Porphyrio hochstetteri*) to island refuges in New Zealand. *Biological Conservation* **94**, 107–114.

Kitaysky A.S., Wingfield J.C. and Piatt J.F. (1999). Dynamics of food availability, body condition and physiological stress response in breeding Black-legged Kittiwakes. *Functional Ecology* **13**, 577–584.

Koller M. (2016). robustlmm: An R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software* **75**, 1–24.

Kruuk L.E., Sheldon B.C. and Merilä J. (2002). Severe inbreeding depression in collared flycatchers (*Ficedula albicollis*). *Proceedings of the Royal Society - London B: Biological Sciences* **269**, 1581–1589.

Kuznetsova A., Brockhoff P.B. and Christensen R.H.B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* **82**, 1–26.

Leuba C., Tebbich S., Nemeth E., Anchundia D., Heyer E., Mosquera D.A. et al. (2020). Effect of an introduced parasite in natural and anthropogenic habitats on the breeding success of the endemic little vermilion flycatcher *Pyrocephalus nanus* in the Galápagos. *Journal of Avian Biology* **51**, e02438.

Lundh J. (2006). The farm area and cultivated plants on Santa Cruz, 1932–1965, with remarks on other parts of Galápagos. *Galapagos Research* **64**, 12–25.

Mahmood I., Imadi S.R., Shazadi K., Gul A. and Hakeem K.R. (2016). Effects of pesticides on environment. In Hakeem K., Akhtar M. and Abdullah S. (eds), *Plant, Soil and Microbes*, vol. 1, *Implications in Crop Science*. Cham: Springer, pp. 253–269.

Mannan R.W. and Meslow E.C. (1984). Bird populations and vegetation characteristics in managed and old-growth forests, northeastern Oregon. *Journal of. Wildlife Management* **48**, 1219–1238.

Mayfield H. (1975). Suggestions for calculating nesting success. *The Wilson Bulletin.* **87**, 456–466.

Meijer T., Masman D. and Daan S. (1989). Energetics of reproduction in female kestrels. *The Auk* **106**, 549–559.

Merlen G. (2011). Gone, gone… going: The fate of the Vermilion Flycatcher on Darwin's Islands. *Galapagos Report 2012*. Puerto Ayora, Galapagos: GNPS, GCREG, CDF and GC, pp. 180–188.

Mosquera D., Fessl B., Anchundia D., Heyer E., Leuba C., Nemeth E. et al. (2022). The invasive parasitic fly *Philornis downsi* is threatening little vermilion flycatchers on the Galápagos Islands. *Avian Conservation and Ecology* **17**, 6.

O'Brien E.L. and Dawson R.D. (2011). Plumage color and food availability affect male reproductive success in a socially monogamous bird. *Behavioral Ecology* **22**, 66–72.

Ortega-Álvarez R. and Lindig-Cisneros R. (2012). Feathering the scene: the effects of ecological restoration on birds and the role birds play in evaluating restoration outcomes. *Ecological Restoration* **30**, 116–127.

Ouyang J.Q., Quetting M. and Hau M. (2012), Corticosterone and brood abandonment in a passerine bird. *Animal Behaviour* **84**, 261–268.

R Core Team (2022), *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Renteria J.L., Atkinson R., Crespo C., Gardener M.R. and Grosholz E.D. (2021), Challenges for the management of the invasive blackberry (*Rubus niveus*) in the restoration of the Scalesia forest in the Galapagos Islands. *Invasive Plant Science and Management* **14**, 20–28.

Renteria J.L., Gardener M.R., Panetta F.D., Atkinson R. and Crawley M.J. (2012), Possible impacts of the invasive plant *Rubus niveus* on the native vegetation of the Scalesia forest in the Galapagos Islands. *PLOS ONE* **7**, e48106.

Rivas-Torres G.F., Benítez F.L., Rueda D., Sevilla C. and Mena C.F. (2018). A methodology for mapping native and invasive vegetation coverage in archipelagos: An example from the Galápagos Islands. *Progress in Physical Geography: Earth and Environment* **42**, 83–111.

Robinson S., Simmons L. and Kennington W. (2014). Sex-biased mortality associated with inbreeding in *Drosophila melanogaster*. *BMC Evolutionary Biology* **14**, 1–9.

Rothschild L.W.R.B. and Hartert E. (1899). A review of the ornithology of the Galapagos Islands: with notes on the Webster-Harris expedition. *Novitates Zoologicae*. **1899**, 85–205.

Rubenstein D.R. (2007). Territory quality drives intraspecific patterns of extrapair paternity. *Behavioral Ecology* **18**, 1058–1064.

Sekercioglu C.H. (2002). Forest fragmentation hits insectivorous birds hard. *Directions in Science* **1**, 62–64.

Tallamy D.W. and Shriver W.G. (2021). Are declines in insects and insectivorous birds related? *Ornithological Applications* **123**, duaa059.

Václav R., Hoi H. and Blomqvist D. (2003). Food supplementation affects extrapair paternity in house sparrows (*Passer domesticus*). *Behavioral Ecology* **14**, 730–735.

Wagner D.L., Grames E.M., Forister M.L., Berenbaum M. and Stopak D. (2021). Insect decline in the Anthropocene: Death by a thousand cuts. *Proceedings of the National Academy of Sciences – PNAS* **118**, e2023989118.

Walentowitz A., Manthey M., Bentet Preciado M.B., Chango R., Sevilla C. and Jäger H. (2021). Limited natural regeneration of unique Scalesia forest following invasive plant removal in Galapagos. *PLOS ONE* **16**, e0258467.

Watson J., Trueman M., Tufet M., Henderson S. and Atkinson R. (2010). Mapping terrestrial anthropogenic degradation on the inhabited islands of the Galapagos Archipelago. *Oryx* **44**, 79–82.

Wauters N., Fournier D. and Dekoninck W. (2017). Do agricultural zones on Santa Cruz Island, Galápagos, host native and endemic arthropods? *The Pan-Pacific Entomologist* **93**, 141–158.

Zeileis A. and Hothorn T. (2002 Diagnostic checking in regression relationships. *R News* **2**, 7–10.

SUPPLEMENTARY INFORMATION

Table S1. Comparison of logistic regression models fitted to daily nest failure data for Little Vermilion Flycatchers. Each row of the table reports the AIC_c value for each of five models fitted to the data (see main text and Table 1). The model specification columns show which of the two independent variables Season and Treatment and their two-way interaction term were included in the model. All models also included the stage of nesting (egg-stage; nestling stage) as an additional binary independent variable (see main text).

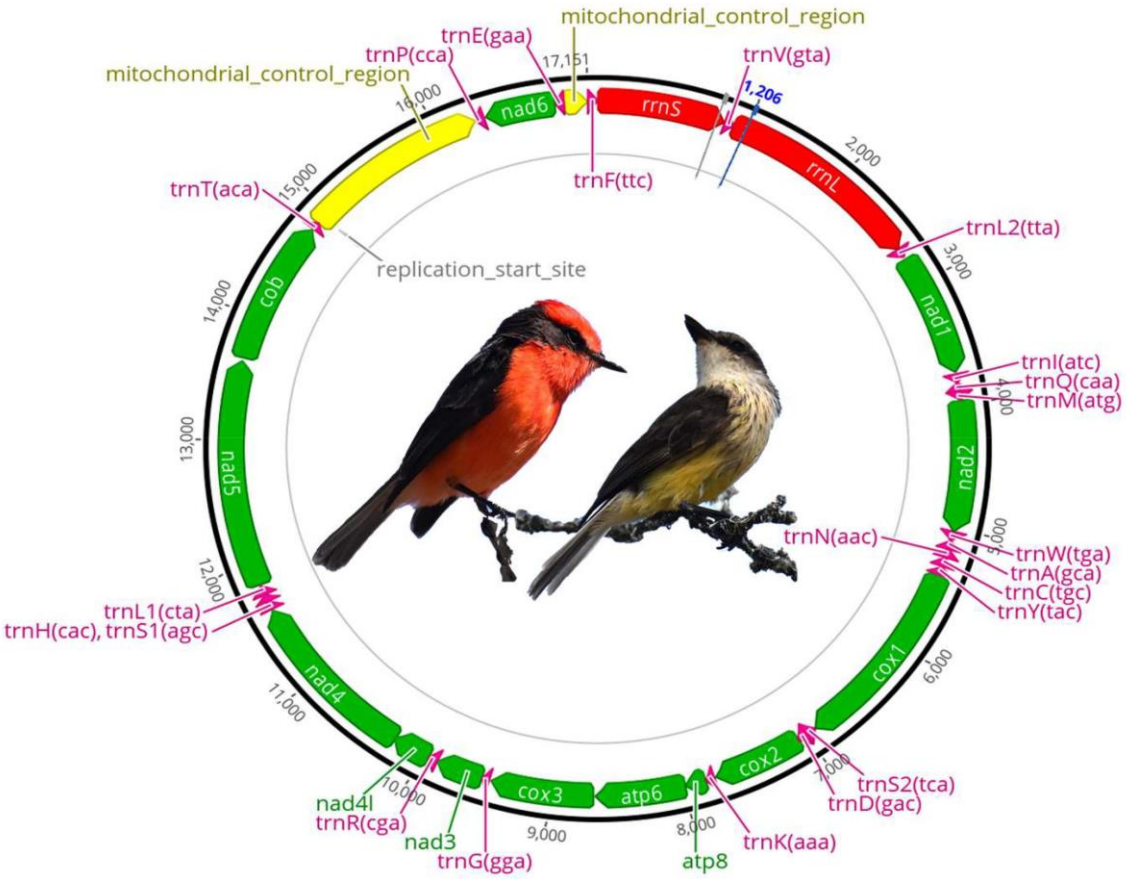
	Model specification				
Model code	Season	Treatment	Season.Treatment	Number of fitted parameters	AIC_c
0	0	0	0	2	87.522
1	1	0	0	3	88.503
2	0	1	0	3	86.278
3	1	1	0	4	88.014
4	1	1	1	5	85.807

Figure S1. a) Female sitting on a cup nest incubating the eggs. The nest is built by the female and comprised of moss, twigs, dry leaves, and sometimes feathers, hair and plant fiber are incorporated, b) nest with nestlings inside, and c) a nest with eggs inside. Photos by David Anchundia and George Gutiérrez.

CHAPTER 2

Genome assembly of *Pyrocephalus nanus*, a step toward the genetic conservation of the endangered Vermilion Flycatchers of the Galapagos Islands

DAVID J. ANCHUNDIA · ATHENA W. LAM · JAMES B. HENDERSON · MATTHEW H. VAN DAM · JOHN P. DUMBACHER



Genome Biology and Evolution (2024) 16(5): evae083. https://doi.org/10.1093/gbe/evae083
Accepted: 13 April 2024 / Published online: 23 April 2024 / Corrected and typeset: 8 May 2024

GBE

Genome Assembly of *Pyrocephalus nanus*: A Step Toward the Genetic Conservation of the Endangered Little Vermilion Flycatcher of the Galapagos Islands

David J. Anchundia [1,2,3,*], Athena W. Lam [3], James B. Henderson[3], Matthew H. Van Dam [3], and John P. Dumbacher [3]

[1]Department of Behavioral and Cognitive Biology, University of Vienna, Vienna 1030, Austria

[2]Charles Darwin Research Station, Charles Darwin Foundation, Santa Cruz, Galapagos, Ecuador

[3]Institute for Biodiversity Sciences and Sustainability, California Academy of Sciences, San Francisco, CA 94118, USA

*Corresponding author: E-mail: davidanchundia@gmail.com.

Accepted: April 13, 2024

Abstract

Incredibly powerful whole genome studies of conservation genetics, evolution, and biogeography become possible for non-model organisms when reference genomes are available. Here, we report the sequence and assembly of the whole genome of the little vermilion flycatcher (*Pyrocephalus nanus*; family Tyrannidae), which is an endemic, endangered, and declining species of the Galapagos Islands. Using PacBio HiFi reads to assemble long contigs and Hi-C reads for scaffolding, we assembled a genome of 1.07 Gb comprising 267 contigs in 152 scaffolds, scaffold N50 74 M, contig N50 17.8 M, with 98.9% assigned to candidate chromosomal sequences and 99.72% of the BUSCO passeriformes 10,844 single-copy orthologs present. In addition, we used the novel HiFiMiTie pipeline to fully assemble and verify all portions of the mitochondrial genome from HiFi reads, obtaining a mitogenome of 17,151 bases, containing 13 protein-coding genes, 22 tRNAs, 2 rRNAs, two control regions, and a unique structure of control region duplication and repeats. These genomes will be a critical tool for much-needed studies of phylogenetics, population genetics, biogeography, and conservation genetics of *Pyrocephalus* and related genera. This genome and other studies that use it will be able to provide recommendations for conservation management, taxonomic improvement, and to understand the evolution and diversification of this genus within the Galapagos Islands.

Key words: *Pyrocephalus nanus*, Tyrannidae genome, Galapagos, vermilion flycatcher.

Significance

The genus *Pyrocephalus* (family Tyrannidae) comprises four recognized species with nine subspecies, distributed within North, Central, and South America, including the Galapagos Islands. The taxonomy of the genus is in flux, and the species endemic to Galapagos include one recently extinct species (*Pyrocephalus dubius*), and several declining populations of another vulnerable species (*Pyrocephalus nanus*). This genome will provide valuable reference for much-needed phylogenetic, population genetic, and conservation genetic work on this genus and species. This and additional studies are being done to help advise active management of this species in Galapagos.

Introduction

The Galapagos Islands are recognized for their geographical isolation, high endemism, and the biogeographic patterns of evolution within the archipelago. Of 213 native bird species recorded, 48 are endemic (Charles Darwin Foundation 2024). Of these endemic species, the complete genome has been assembled at the scaffold level for only four species, including the Medium Ground Finch (*Geospiza fortis*) (Zhang et al. 2014), the Galapagos Flightless Cormorant (*Phalacrocorax harrisi*) (Burga et al. 2017), Galapagos Penguin (*Spheniscus mendiculus*) (Pan et al. 2019), and the Small Tree Finch (*Camarhynchus parvulus*) (Rubin et al. 2022).

Due to increasing anthropogenic pressure and decreasing population sizes of many endemic birds in Galapagos (Fessl et al. 2015; Dvorak et al. 2017, 2020; Geladi et al. 2021), there is a pressing need for high-resolution genetic conservation studies of threatened species in Galapagos. One important endemic species is the Galapagos little vermilion flycatcher, *Pyrocephalus nanus*, which was once distributed throughout most of the Galapagos archipelago (12 islands) (Gifford 1919), but in recent decades, its populations have disappeared from four islands and are on the verge of disappearing from two more (Merlen 2013; Fessl et al. 2015; Leuba et al. 2020).

Pyrocephalus nanus is found in multiple distinct climate and ecosystem types (BirdLife International 2023) from xeric/desert ecosystems near the coast (Rothschild and Hartert 1899; Gifford 1919) to the tops of volcanoes with evergreen humid vegetation and frequent precipitation (Mosquera et al. 2022). The lack of a reference genome or detailed genetic study has limited investigations into its population history, taxonomy, evolutionary biology, and its adaptations to various ecosystem types, although preliminary work suggests that there is significant variation among populations (Carmi et al. 2016). Current bird conservation programs in Galapagos aim to recover and restore the former range distribution of this species, and therefore, knowledge obtained from genomic data will be useful for conservation planning and action (García-Dorado and Caballero 2021). The aim of this study is to assemble the whole genome of *P. nanus* as a tool for guiding research and management. Additionally, the genome should be a useful resource for studies of one of the largest avian families Tyrannidae, which includes *Pyrocephalus*.

Results

Reference Whole Genome

We sequenced two cells of PacBio HiFi reads that produced 6.84 million reads with 788.79 billion bases, N50 read length 12,741 bp and mean read length 12,551 bp. The Hi-C library produced 103.42 million paired reads, and we retained 99.94 million paired reads after cleaning. We removed 145 sequences of 19,364,419 bases from the HiFiasm contig-only assembly classified by purge_dups as Haplotig, Repeat, Junk, or Highcov. After contig assembly, removal of purge_dup sequences, scaffolding with Hi-C reads, and manual contig placement and orientation, we obtained a final assembly of 1.073 Gb, in 152 scaffolds and 267 contigs (Fig. 1 and Table 1). By comparing the long scaffolds to the chromosome-level assembled genome of the zebra finch (NCBI RefSeq GCF_003957565.2), and using telomere sequences at the end or beginning, we were able to reorganize scaffolds of the *P. nanus* genome into 38 chromosomes (supplementary table S2). Additionally, we were able to identify the fifth largest chromosome as sex chromosome Z (Fig. 1) in addition to the 37 autosomes. Of the 38 chromosome designated sequences, 34 contain at least one telomere and 10 of them have a top and a bottom telomere.

Assessment of the Nuclear Genome

GenomeScope2 kmer 21 modeling showed a smaller genome at 791 Mbp than the 1.07 Gbp final assembly size, and 2% heterozygosity with only 45 Mbp repeats. HiFiasm showed kmer peak homozygosity of 24, matching GenomeScope's 24.1 assessment, and did not find a heterozygosity peak. The genome size of 1.07 Gbp had repeats of 120 M bases, 11.20%, which is often underrepresented in the GenomeScope model. To assess the completeness of the genome, we used BUSCO v.5.4.7 (Simão et al. 2015) and compleasm v0.2.2 (Huang and Li 2023) and combined results. BUSCO + compleasm identified Complete BUSCO genes: 99.72% (single-copy orthologs: 99.63%, duplicated: 0.09%, fragmented: 0.06%, missing: 0.22%, *n*: 10,844 genes) (Table 1, see Supplementary Material, part II).

Genome Annotation

Repeatmasker masked 11.20% (see supplementary table S1) of the 1.07 Gbp genome, or 120 Mbp as repeats. This is less than many other avian genomes but similar to several other flycatcher genomes recently sequenced. All annotations were done bioinformatically without RNA-seq data. BRAKER initially found 31,920 start codons and 31,931 stop codons. Functional gene annotation was used to remove low confidence gene models. The final gene annotation included 30,101 genes and 31,748 mRNAs, with a total of 161,303 exons. Genes composed 13.26% of the genome. Gene names and/or descriptions were assigned to 19,303 of the identified mRNAs from the genome. Additional annotation statistics can be found in the supplement, as well as full annotation (.gff, .fna, .faa) files (included as Supplementary Material).

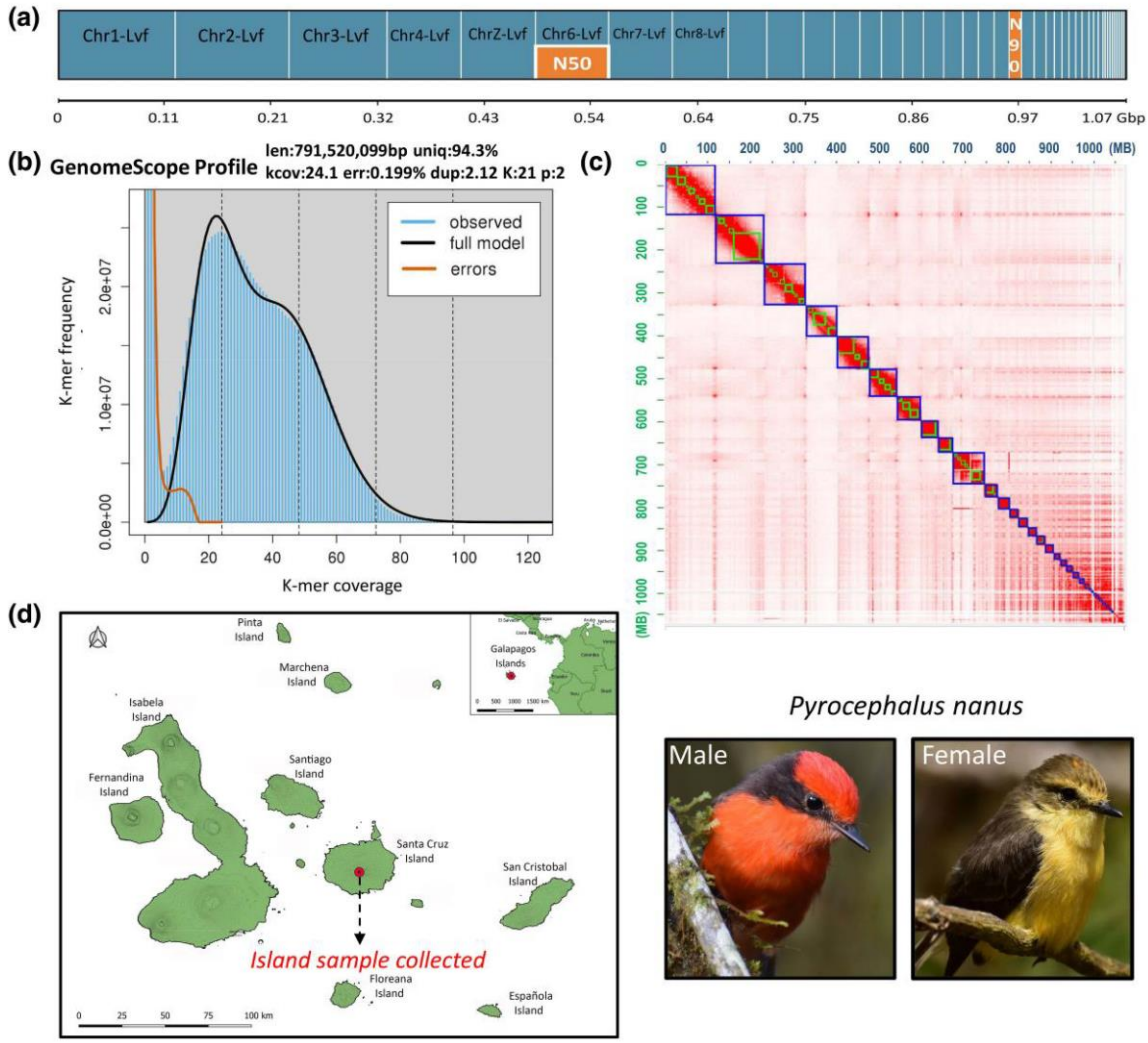


Fig. 1.—a) Graphical representation of assembled genome statistics, genome size 1.07 Gigabases (Gbp) with N50 and N90 scaffold size comparison graph produced with Quast (Gurevich et al. 2013). Chr 1,2,3… means the chromosome and its number located in the assembled genome of LVF (little vermilion flycatcher), ChrZ-Lvf is the sex chromosome. b) kmer histogram produced with GenomeScope2 to see genome length (len) derived from PacBio sequences and kmer coverage (kcov), read error rate (err), genome unique length (uniq), and duplicate (dup). c) Hi-C contact maps from Juicebox Assembly Tools (JBAT) v.1.11.08 after manual refinement, orientation of contigs, and their delineation within chromosomes, scale in megabases (Mb). d) *Pyrocephalus nanus* is endemic to the Galapagos archipelago, a volcanic island system located in the Eastern Tropical Pacific. A red dot in the center of the map highlights the highlands of Santa Cruz Island *where* the sample was collected. The birds are sexually dimorphic, on the right, you can see the adult plumage.

Mitochondrial Genome

A total of 167 HiFiasm corrected HiFi reads mapped to the mitogenome, and the HiFiMiTie pipeline unambiguously assembled these into a complete circular mitochondrial genome of 17,151 nucleotides. The genome has a remnant control region (CR) flanked by tRNA Glu and tRNA Phe, and a primary control region flanked by tRNA Thr and tRNA Pro (see supplementary fig. S1, Supplementary Material online, Supplementary Material). A total of 113 HiFi reads contained control region segments: each of the remnant CR sections was 178 nt in length, and each of the primary CR sections was 1,430 nt in length. As typical for avian mitogenomes, there were 22 tRNAs, 2 rRNAs, and 13 protein-coding genes (supplementary fig. S1, Supplementary Material online; see details in part I of Supplementary Material).

As a check, the pipeline runs megahit (Li et al. 2016) using the same 167 HiFi reads. The resulting sequence is identical except where the primary control region contains repeats. The issue with repeat contraction or duplication is a common problem with kmer based assemblers such as megahit. The HiFiMiTie primary assembly mode uses segmented multiple sequence alignment and is typically a

Table 1
Sample information assembled genome statistics and accession numbers

Species Scientific name	*Pyrocephalus nanus*	Common name	Little vermilion flycatcher			
Location sample collection	Island: Santa Cruz, Province: Galapagos Islands, Country: Ecuador	Coordinate	S0.63172° W90.36300°			
BioProjects and voucher	NCBI BioProject	PRJNA1040305				
	NCBI BioSample	SAMN38255597				
	NCBI Genome accession	JAWZSU000000000				
	Genome assembly	bPyrNan1_0.fasta				
	NCBI SRA accession raw reads data	PRJNA1040305				
Genome sequence	PacBio HiFi reads	Run	2 PacBio SMRT cells 6.84 million HiFi reads			
	Illumina HiC	Run	NovaSeq 6000 103.42 million pair reads			
Genome assembly quality metrics	Number of scaffolds	152				
	Total size of scaffolds	1,072,479,546				
	Longest scaffold	117,695,848				
	Shortest scaffold	12,661				
	Mean scaffold size	7,055,786				
	Median scaffold size	81,135				
	N50 scaffold length	74,038,366				
	N90 scaffold length	12,777,097				
	L50 scaffold count	6				
	L90 scaffold count	21				
	N50 contig length	17,811,915				
	N90 contig length	4,626,274				
	Number of contigs	267				
	Number of contigs in scaffolds	144				
	Number of contigs not in scaffolds	123				
	GC (%)	42.30				
	Gaps 100 Ns	115				
	BUSCO completeness passeriformes	C	S	D	F	M
	$n = 10,844$	99.7%	99.6%	0.1%	0.1%	0.2%
	Organelle	Whole mitochondrial genome 17,151 bp	22 tRNAs, 2 rRNAs, 2 CRs and 13 PCGs			

more reliable tool for resolving repeated regions when highly accurate long reads are available.

Discussion

Here, we present a highly complete genome of *P. nanus*, and the first genome of a Galapagos passerine assembled with PacBio HiFi sequences. This haploid genome size of 1.073 Gb is similar to other genomes of this bird family that are between 1.0 and 1.1 Gb (Ruegg et al. 2018). Of the eight complete genomes currently assembled on the Tyrannidae family (NCBI), none were made with PacBio HiFi sequences. These other assemblies have a large number of scaffolds ranging from 1,692 to 43,947 with a mean of 16,422 see supplementary table S3, Supplementary Material online, Supplementary Material. Using HiFi reads, we were able to obtain longer average contig sizes and with Hi-C reads fewer scaffolds. This high-resolution genome is also an important resource for future studies on the Tyrannidae family (441 species), the largest avian family (Billerman et al. 2020), and the subfamily Fluvicolinae (~130 species) that includes *Pyrocephalus*, its closest related taxa *Alectrurus*, *Arundicola*, *Gubernetes*, *Colonia*, and several other poorly studied genera (Ohlson et al. 2008; Feng et al. 2020). Also, in our haplotype assembly, we obtained a BUSCO + compleasm complete score of 99.72% suggesting that a great conservation of genes in our assembly and that the genome assembly is highly complete (Huang and Li 2023). This assembly is a step forward to conduct more genomic studies on the endemic species of the Galapagos archipelago.

Materials and Methods

Tissue Collection

A single egg of *P. nanus* was collected while monitoring nests on Santa Cruz Island during the 2021 breeding season. The nest was blown to the ground by strong winds where the author collected a single broken egg containing a significantly developed but dead embryo. Approximately 40 min after death, the embryo was preserved in 96% alcohol and stored in a freezer at −27 °C. Prior to analysis, the

frozen sample was exported from Galapagos to the California Academy of Sciences.

PacBio High-Molecular-Weight DNA Extraction, Library Preparation, and Sequencing

We prepared libraries using standard PacBio recommendations, but full protocol details for library prep can be found in PacBio (2012) and Van Dam et al. (2021). We confirmed large amounts of high-molecular-weight DNA using FemtoPulse (Agilent, Santa Clara, CA). DNA was sent to the QB3 Genomics facility at the University of California Berkeley for HiFi library preparation and sequencing on a Pacific Biosciences Sequel II platform, sequencing two SMRT cells with HiFi version 2 chemistry.

In Situ Hi-C Library Preparation

Additional muscle tissues from the same sample were homogenized using a sterile razor blade on ice. An in situ Hi-C library was prepared as described in Rao (2014) with a few modifications. Briefly, after the streptavidin pull-down step, the biotinylated Hi-C products underwent end repair, ligation, and enrichment steps using the NEBNext UltraII DNA Library Preparation kit (New England Biolabs Inc, Ipswich, MA). Titration of the number of PCR cycles was performed as described in Belton (Belton et al. 2012).

Genome Assembly

For HiFi data preparation, cutadapt v.4.4 (Martin 2011) was used to remove any read with length less than 1000 bp or that contained a PacBio SMRTbell adapter in any position. Cutadapt arguments revcomp, error-rate 0.1, overlap 35, discard, minimum-length 1000 were used along with the –b adapter argument to create cleaned fastq HiFi reads. To assess genome size, we ran jellyfish v2.3.0 (Marçais and Kingsford 2011) using its count option with long reads and kmer size 21, then jellyfish histo for a histogram of kmer frequencies. The histo file was uploaded to GenomeScope2 (qb.cshl.edu/genomescope/genomescope2.0) (Ranallo-Benavidez et al. 2020) to provide estimates of genome properties including total size, repeat content, and heterozygosity.

The two PacBio HiFi cleaned fastq read sets were assembled into genome contigs using the program hifiasm v.0.16.1-r375 (Cheng et al. 2021) and run with arguments —write-ec —write-paf. The HiFiasm program was run via a custom script that converts the program's gfa output to fasta files with any circular records stored separately. Various statistics files were created from the fasta file, including N50, N90, and telomere location in contigs, and BUSCO v.5.4.7 (Simão et al. 2015) was run using the avian passeriformes lineage dataset. To remove haplotypic duplicates, we ran purge_dups v.1.2.5 (Guan et al. 2020) using cutoffs -l 5 -m 36 -u 108 and excluded records from the contig assembly identified as duplicative. To scan for contaminants, we used Kraken2 (Wood et al. 2019) and blastn (Camacho et al. 2009). Taxonomy ID results from the blastn search were translated and sorted by clade, allowing for the identification of any non-avian contigs that were then removed from the contig-level assembly.

To scaffold contigs using Hi-C reads, the Illumina reads from the Hi-C tissue were cleaned and prepared in two steps. First, fastp v.0.23.2 (Chen et al. 2018) was run with the dedup argument to remove Illumina adapters and any read less than 100 bp, and its pair, after adapter removal. Following this, Arima's pipeline (github.com/ArimaGenomics/mapping_pipeline) (Arima 2021) was used with the fastp cleaned Hi-C reads as input to perform additional clean-up and to map the reads to the contig-level assembly with bwa (Li and Durbin 2009; Li 2013). The resulting bam file and the contig assembly were input into the YaHS v.1.2 scaffolding program (Zhou et al. 2023). YaHS scaffolding was run with bam file mapped Hi-C reads and fasta contig assembly input and the –no-contig-ec option. YaHS created .hic and .assembly files that were used to display Hi-C contact maps in Juicebox Assembly Tools (JBAT) v.1.11.08 (Durand et al. 2016a, 2016b) for manual refinement, and we interactively updated the location and orientation of contigs and their delineation within chromosomes.

To assess the level of genome completeness, we ran both compleasm v.0.2.2 (Huang and Li 2023; a reimplementation of BUSCO using miniprot (Li 2023)), and BUSCO v.5.4.7 (Simão et al. 2015) with its default MetaEuk (Levy-Karin et al. 2020) mode, each using the 10,844 ortholog lineage dataset passeriformes_odb10. We updated any BUSCOs not found by compleasm that were found in the BUSCO MetaEuk results.

Genome Annotation

Prior to gene annotation, regions with repeats were identified using RepeatModeler v2.0.1 (Flynn et al. 2020). We combined repeat models found using RepeatModeler with the avian repeat models from Repbase RepeatMasker libraries v20181026 into a single fasta file. These combined repeat models were used in Repeatmasker v4.0.9 (Smit et al. 2015) with the options -small -xsmall and -nolow to create a soft-masked repeat version of the assembly file used for the gene model structural annotation as well as to create a table of repeat types and lengths.

We annotated the genome without RNA sequence data using BRAKER (version 3.03 April 2023) to predict the location of genes (mRNAs, introns, exons, CDS) using BRAKER's ProtHint pipeline (version 2.6.0 Georgia Tech GeneMark) and AUGUSTUS 3.5.0 (Gabriel et al. 2023) with the vertebrate amino acid sequences from Vertebrata_OrthoDB_10 (Kuznetsov et al. 2023) BRAKER pipeline B., previously called

BRAKER2, which is employed when RNA-seq data is unavailable. Potential gene positions were output to gff files. We then eliminated any potential genes that didn't have both a start and a stop codon and genes that were fully nesting within other genes. Functional annotation of these genes was begun by looking for protein domains in the amino acid sequences found by BRAKER by running InterProScan-5.61-93.0 (with CDD-3.20, FunFam-4.3.0, PANTHER-17.0, Pfam-35.0, PIRSF-3.10, PRINTS-42.0, ProSitePatterns-2022_05, ProSiteProfiles-2022_05, S MART-9.0, SUPERFAMILY-1.75, TIGRFAM-15.0 analyses). The sequences were also blasted against Genbank nt, nr, and swissprot databases downloaded on March 27, 2023 and UNIPROT TrEMBL downloaded on May 15, 2021. Coding sequences were blasted using BLASTN (v2.14) using the nt database. Translated protein sequences were blasted using BLASTP (v2.14) with SwissProt, diamond blastp v.2.1.6 (Buchfink et al. 2021) with the TrEMBL database and with OrthoDB10 vertebrate orthologs. Protein domain IDs and Gene Ontology terms, from InterProScan output, were added to the gff file for each gene and isoform model as were the functional annotation description from the lowest eValue, highest score result from the blast searches. They were also added to the Amino Acid and to the CDS fasta file gene sequences.

Mitochondrion Assembly

Mitochondrial sequence was derived from corrected HiFi reads from the nuclear genome assembly using an internally created program pipeline named HiFiMiTie (Henderson 2021) version 0.1. The HiFiMiTie pipeline was designed to extract and assemble mitochondrial sequence from PacBio HiFi long reads and also resolve control region heteroplasmy and repeats. It also discovers and annotates tRNAs, rRNAs, protein-coding genes, and up to two duplicate CRs. See details and full logged output in the Supplementary Material, part I.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

Ecuador is a signatory to the Nagoya protocol; collecting and export permits were approved by the Ministry of Environment of Ecuador and the Galapagos National Park. This study was conducted with permission of the Galapagos National Park Directorate research permit PC-06-21. The Ministry of Environment of Ecuador approved the genetic permit (MAE-DNB-CM-2016-0043) and sample export permit (N° 012-2021-EXP-CM-FAU-DBI/MAAE). This publication is contribution number 2550 of the Charles Darwin Foundation for the Galapagos Islands. The study was funded by a PhD fellowship awarded to D.J.A. from the World Wildlife Fund, Mr. Russell E. Train and Education for Nature, Agreement #RH94, a Lakeside Fellowship to D.J.A. from the California Academy of Sciences, Peter and Kris Norvig donation to the Charles Darwin Foundation, and a donation from Michael Simon to the California Academy of Sciences. Other funding sources to conduct fieldwork include Galapagos Conservation Trust; Galapagos Conservancy and Lindblad Expeditions—National Geographic. We thank Courtney Pike for feedback in improving the manuscript and sample collection.

Data Availability

The assembled genome is available in NCBI with the reference number submission ID SUB13890500 and the accession number JAWZSU000000000. The raw sequence files used in the assembly are available in the NCBI Sequence Read Archive, accession number PRJNA1040305. Scripts and codes used to perform these analyses are on GitHub (https://github.com/calacademy-research/).

Literature Cited

Arima X. Arima genomics mapping pipeline. 2021 [accessed 2019 Feb 9]. https://www.github.com/ArimaGenomics/mapping_pipeline/blob/master/arima_mapping_pipeline.sh.

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods. 2012:58(3):268–276. https://doi.org/10.1016/j.ymeth.2012.05.001.

Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS. Birds of the world. Ithaca, NY: Cornell Laboratory of Ornithology; 2020.

BirdLife International. Pyrocephalus nanus. The IUCN Red List of Threatened Species 2023: e.T103682926A172654604. https://doi.org/10.2305/IUCN.UK.2023-1.RLTS.T103682926A172654604.en

Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021:18(4):366–368. https://doi.org/10.1038/s41592-021-01101-x.

Burga A, Wang W, Ben-David E, Wolf PC, Ramey AM, Verdugo C, Lyons K, Parker PG, Kruglyak L. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. Science. 2017:356(6341):eaal3345. https://doi.org/10.1126/science.aal3345.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009:10(1):1–9. https://doi.org/10.1186/1471-2105-10-421.

Carmi O, Witt CC, Jaramillo A, Dumbacher JP. Phylogeography of the vermilion flycatcher species complex: multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. Mol Phylogenet Evol. 2016:102:152–173. https://doi.org/10.1016/j.ympev.2016.05.029.

Charles Darwin Foundation. Galapagos Species Database, dataZone. Charles Darwin Foundation. 2024 [accessed 2024 May 1]. https://datazone.darwinfoundation.org/en/checklist/.

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018:34(17):884–890. https://doi.org/10.1093/bioinformatics/bty560.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18(2):170–175. https://doi.org/10.1038/s41592-020-01056-5.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016a:3(1):99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016b:3(1):95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Dvorak M, Fessl B, Nemeth E, Anchundia D, Cotín J, Schulze CH, Tapia W, Wendelin B. Survival and extinction of breeding landbirds on San Cristóbal, a highly degraded island in the Galápagos. Bird Conserv Int. 2020:30(3):381–395. https://doi.org/10.1017/S0959270919000285.

Dvorak M, Nemeth E, Wendelin B, Herrera P, Mosquera D, Anchundia D, Sevilla C, Tebbich S, Fessl B. Conservation status of landbirds on Floreana: the smallest inhabited Galapagos Island. J Field Ornithol. 2017:88(2):132–145. https://doi.org/10.1111/jofo.12197.

Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, Xie D, Chen G, Guo C, Faircloth BC, et al. Dense sampling of bird diversity increases power of comparative genomics. Nature. 2020:587(7833):252–257. https://doi.org/10.1038/s41586-020-2873-9.

Fessl B, Anchundia D, Carrion-Tacuri J, Cimadom A, Cotın J, Cunninghame F, Dvorak M, Mosquera D, Nemeth E, Sevilla CR, et al. Galapagos landbirds (passerines, cuckoos, and doves): Status, threats and knowledge gaps. In Galapagos Reports 2015-2016. GNPS, GCREG, CDF and GC. Puerto Ayora, Galapagos, Ecuador. 2015.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117(17):9451–9457. https://doi.org/10.1073/pnas.1921046117.

Gabriel L, Brna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. bioRxiv. 2023. preprint this version posted February 29, 2024. https://doi.org/10.1101/2023.06.10.544449.

García-Dorado A, Caballero A. Neutral genetic diversity as a useful tool for conservation biology. Conserv Genet. 2021:22(4):541–545. https://doi.org/10.1007/s10592-021-01384-9.

Geladi I, Henry P-Y, Mauchamp A, Couenberg P, Fessl B. Conserving Galapagos landbirds in agricultural landscapes: forest patches of native trees needed to increase landbird diversity and abundance. Biodivers Conserv. 2021:30(7):2181–2206. https://doi.org/10.1007/s10531-021-02193-9.

Gifford EW. Expedition of the California Academy of Sciences to the Galapagos Islands, 1905–1906. XIII. Field notes on the land birds of the Galapagos Islands and of Cocos Island, Costa Rica. Proc Cal Acad Sci Fourth Series. 1919:2:189–258.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020:36(9):2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013:29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Henderson J. HiFiMiTie find & analyze metazoan mitochondria from HiFi reads. 2021 California Academy of Sciences, Institute for Biodiversity Science & Sustainability. [accessed 2021 Nov 15]. https://github.com/calacademy-research/HiFiMiTie.

Huang N, Li H. compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics. 2023:39(10):btad595. https://doi.org/10.1093/bioinformatics/btad595.

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 2023:51(D1):D445–D451. https://doi.org/10.1093/nar/gkac998.

Leuba C, Tebbich S, Nemeth E, Anchundia D, Heyer E, Mosquera DA, Richner H, Rojas Allieri ML, Sevilla C, Fessl B. Effect of an introduced parasite in natural and anthropogenic habitats on the breeding success of the endemic little vermilion flycatcher *Pyrocephalus nanus* in the Galápagos. J Avian Biol. 2020:51(8):1–13. https://doi.org/10.1111/jav.02438.

Levy-Karin E, Mirdita M, Söding J. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. Microbiome. 2020:8(1):1–15. https://doi.org/10.1186/s40168-020-00808-x.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303-3997, 26 May 2013, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1303.3997.

Li H. Protein-to-genome alignment with miniprot. Bioinformatics. 2023:39(1):btad014. https://doi.org/10.1093/bioinformatics/btad014.

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009:25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016:102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011:27(6):764–770. https://doi.org/10.1093/bioinformatics/btr011.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011:17(1):10–12. https://doi.org/10.14806/ej.17.1.200.

Merlen G. Gone, gone…going: the fate of the vermilion flycatcher on Darwin's islands. Galapagos Report. 2013:2012:180–188.

Mosquera D, Fessl B, Anchundia D, Heyer E, Leuba C, Nemeth E, Rojas Allieri ML, Sevilla CR, Tebbich S. The invasive parasitic fly Philornis downsi is threatening little vermilion flycatchers on the Galápagos Islands. Avian Conserv Ecol. 2022:17(1):6. https://doi.org/10.5751/ACE-02040-170106.

National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]—[cited 2023 Aug 30]. https://www.ncbi.nlm.nih.gov/assembly/?term=tyrannidae.

National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]—[cited 2023 Oct 24]. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_003957565.2/

Ohlson J, Fjeldså J, Ericson PG. Tyrant flycatchers coming out in the open: phylogeny and ecological radiation of Tyrannidae (Aves, Passeriformes). Zool Scr. 2008:37(3):315–335. https://doi.org/10.1111/j.1463-6409.2008.00325.x.

Pacbio Extracting DNA Using Phenol-Chloroform. [cited 25 July 2023]. Available: https://www.pacb.com/wp-content/uploads/2015/09/SharedProtocol-Extracting-DNA-usinig-Phenol-Chloroform.pdf. 2012.

Pan H, Cole TL, Bi X, Fang M, Zhou C, Yang Z, Ksepka DT, Hart T, Bouzat JL, Argilla LS, et al. High-coverage genomes to elucidate the evolution of penguins. GigaScience. 2019:8(9):giz117. https://doi.org/10.1093/gigascience/giz117.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11(1):1432. https://doi.org/10.1038/s41467-020-14998-3.

Rao SS. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014:159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

Rothschild LWRB, Hartert E. A review of the ornithology of the Galápagos Islands: with notes on the Webster-Harris expedition. Zoological Museum. 1899:85–142.

Rubin CJ, Enbody ED, Dobreva MP, Abzhanov A, Davis BW, Lamichhaney S, Pettersson M, Sendell-Price AT, Sprehn CG, Valle CA, et al. Rapid adaptive radiation of Darwin's finches depends on ancestral genetic modules. Sci Adv. 2022:8(27):eabm5982. https://doi.org/10.1126/sciadv.abm5982.

Ruegg K, Bay RA, Anderson EC, Saracco JF, Harrigan RJ, Whitfield M, Paxton EH, Smith TB. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. Ecol Lett. 2018:21(7):1085–1096. https://doi.org/10.1111/ele.12977.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. 2015:289–300.

Van Dam MH, Cabras AA, Henderson JB, Rominger AJ, Pérez Estrada C, Omer AD, Dudchenko O, Lieberman Aiden E, Lam AW. The Easter Egg Weevil (*Pachyrhynchus*) genome reveals syntenic patterns in Coleoptera across 200 million years of evolution. PLoS Genet. 2021:17(8):e1009745. https://doi.org/10.1371/journal.pgen.1009745.

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019:20(1):1–13. https://doi.org/10.1186/s13059-019-1891-0.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014:346(6215):1311–1320. https://doi.org/10.1126/science.1251385.

Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023:39(1):btac808. https://doi.org/10.1093/bioinformatics/btac808.

Associate editor: Qi Zhou

Supplementary Material

Part I. Mitochondrial Genome Assembly using the HiFiMiTie pipeline.

The HiFiMiTie (hereafter called hfmt) workflow uses known mitochondrial genomes from closely-related taxa (in this case Aves) and blastn (version 2.12.0 with arguments -db mito -query $reads -outfmt "6 std staxid stitle qlen qcovhsp qcovus" -max_target_seqs 5 -taxidlist aves_taxids.txt -evalue 1e-10 -mt_mode 1) to search the taxonomic id matching entries in a copy of the NCBI GenBank mitochondrial database, using HiFi reads as queries. Reads with qcovus 50% or greater are extracted and reverse complemented if necessary to match the subject mitochondrion direction. Candidate reads are annotated with the mitochondrial tRNA finder MiTFi (Allio et al. 2020) using the mito genetic code associated with its taxon, here Aves. Additionally, 12S and 16S rRNAs are discovered using Infernal's cmsearch. Control region goose hairpins and any OH, OL origins of replication are also annotated. The right neighbor tRNA of tRNAs in each candidate long read are counted and totals used to create the canonical tRNA order of the mitochondrion. Reads with this order for the subset of tRNAs in them are chosen for all downstream sequence creation; though the annotations of all candidate reads are available for additional analysis and corroboration. The full documentation and files are available from github: https://github.com/calacademy-research/HiFiMiTie.

According to convention for vertebrate mitochondrial genomes, the tRNA succeeding the Control Region is chosen as the start of the mitochondrial genome sequence, which is Phenylalanine (tPhe or F) for our genome as well as most vertebrates. This plus the preceding tRNA from the derived canonical order are designated as a first Control Region. Annotated goose hairpins that occur elsewhere in the HiFi reads are used to identify second control regions. When additional goose hairpins are found and one of the control regions is below a threshold size it is identified as a remnant and the other as the primary, which is the case with our data.

After a primary control region is found, three sets of reads are created and each used as input to mafft (Katoh and Standley 2013) to form multiple sequence alignments (msa). The first set is extracted from reads with the tRNA after the CR up to the CR or the end of the read if no CR is after the tRNA. Second set extracted from reads from the beginning of a read up to the tRNA or rRNA that precedes the primary CR. And the third from those reads with both CR flanking tRNAs with the CR in between.

From the mafft msa alignment of the first 2 sets, a consensus sequence is created from each, and the two are combined to create a mitochondrion sequence excluding the CR. The CR reads are analyzed

for heteroplasmy, typically different copy numbers of repeats, if any, and representative versions are chosen for additional downstream analysis. The most common CR is used to complete the mitochondrion sequence. The sequence is constructed to start at tRNA tPhe, which canonically begins annotations of vertebrate mitogenomes.

The mitochondrial sequence is annotated by the hfmt pipeline. Separately MITOS2 (Bernt et al. 2013) was used to confirm and check the annotation. When running HFMT, the program creates an extensive log file that is well annotated for novice users. The complete log file for our mitochondrial assembly is pasted below, and it contains descriptions of the results and output of each step. A settings file records a range of values.

hfmt.log:

[2021-11-15 15:28:43] HiFiMiTie version 0.01 -- Find & Analyze Metazoan Mitochondria from HiFi reads

[2021-11-15 15:28:43] Step 1 -- Setup: taxid and HiFi file(s) to use
[2021-11-15 15:28:44] HiFiMiTie directory hfmt_111521 created
[2021-11-15 15:28:44] HiFi file(s) to search for mitochondrial reads:
[2021-11-15 15:28:44] /home3/jdumbacher/pyrocephalus/asm/hifiasm_tst1/vf_asm.ec.fa
[2021-11-15 15:28:52] Aves mitogenomes, taxid 8782, chosen as search targets. 964 Aves mitogenomes in the /ccg/blastdbs/mito db.

[2021-11-15 15:29:04] Step 2 -- blast_to_mito
[2021-11-15 15:29:04] created hifi_mito_matches to hold the blast output tsv files
[2021-11-15 15:29:07] blastn -db /ccg/blastdbs/mito -query /home3/jdumbacher/pyrocephalus/asm/hifiasm_tst1/vf_asm.ec.fa -outfmt "6 std staxid stitle qlen qcovhsp qcovus" -max_target_seqs 5 -num_threads 32 -taxidlist taxidlist -evalue 1e-10 -mt_mode 1
[2021-11-15 15:43:03] completed in 0h13m56s with completion code 0
[2021-11-15 15:43:03] 1 blasts completed in 0h13m56s
[2021-11-15 15:43:15] top accn info: 2 mito code of NC_007975 183540 | Cnemotriccus fuscatus | cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Sauropsida; Sauria; Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Passeriformes; Tyrannidae; Cnemotriccus; |
[2021-11-15 15:43:15] mito code chosen: 2
[2021-11-15 15:43:16] Step 2 completed in 0h14m11s

[2021-11-15 15:43:16] Step 3 -- pull_fofn_cand_recs
[2021-11-15 15:43:16] pulling records from /home3/jdumbacher/pyrocephalus/asm/hifiasm_tst1/vf_asm.ec.fa
[2021-11-15 15:46:00] pulled 167 records, excluded 3 with low coverage.
[2021-11-15 15:46:00] created mito_hifi_rec_db to hold a blast database containing mito candidate fasta records
[2021-11-15 15:46:02] Step 3 completed in 0h2m45s

[2021-11-15 15:46:02] Step 4 -- select_mito_features
[2021-11-15 15:46:02] features being retrieved from NC_007975
[2021-11-15 15:46:02] retrieving mito features for NC_007975
[2021-11-15 15:46:02] mito_analyze.py -rec -q NC_007975 -nh | sort -k3,3n | bioawk_cas {print ">" $2"_"$5, $3, $4, fldcat(6,NF); print $1} | fold -w 120
[2021-11-15 15:46:04] top_match_feature_sequences.fasta file created
[2021-11-15 15:46:04] Step 4 completed in 0h0m02s

[2021-11-15 15:46:05] Step 5 -- blast_features
[2021-11-15 15:46:05] created blast_results to hold the blast output tsv files
[2021-11-15 15:46:05] created cr_analysis to hold control region(s) analysis files
[2021-11-15 15:46:05] blastn -db hfmt_111521/mito_hifi_rec_db/mito_hifi_recs -query hfmt_111521/top_match_feature_sequences.fasta -outfmt "6 std staxid stitle qlen qcovhsp qcovus" -num_threads 8 -task blastn -evalue 1e-10 -max_target_seqs 167 | sort -k2,2V -k9,9n -k12,12nr
[2021-11-15 15:46:06] blastn completed in 0h0m01s results in hfmt_111521/blast_results/mito_feature_match_to_cand_recs.tsv

[2021-11-15 15:46:06] blastn -db OH.fas -query hfmt_111521/mito_hifi_recs.fasta -outfmt "6 std staxid stitle qlen qcovhsp qcovus" -max_target_seqs 5 -subject_besthit -evalue .00001 -num_threads 32
[2021-11-15 15:46:09] both feature set and taxname Aves have starting trna of F, setting first trna to F
[2021-11-15 15:46:10] Step 5 completed in 0h0m05s

[2021-11-15 15:46:10] Step 6 -- rna_search
[2021-11-15 15:46:10] cm_results directory created
[2021-11-15 15:46:10] input files checked
[2021-11-15 15:46:10] search for 12S_s-rna
[2021-11-15 15:47:13] search for 16S_l-rna
[2021-11-15 15:51:31] rrna_rrnS.tbl rrna_rrnL.tbl created in cm_results
[2021-11-15 15:51:31] tRNA search mito reads using: MiTFi - mitochondrial tRNA finder
[2021-11-15 16:13:13] mito_hifi_recs.mitfi created in cm_results
[2021-11-15 16:13:14] trna_right_neighbor.matrix created using [1;34mF[0m as the first tRNA.
[2021-11-15 16:13:14] trna order: F V L2 I Q M W A N C Y S2 D K G R H S1 L1 T P E
[2021-11-15 16:13:14] mito_hifi_recs.cm_anno created with mitfi, goose_hairpin, 12S_rna and 16S_rna cm results
[2021-11-15 16:13:14] cm_anno_right_neighbor.matrix created using [1;34mF[0m as the first tRNA.
[2021-11-15 16:13:14] cm_anno rna order: F 12S V 16S L2 I Q M W A N C Y S2 D K G R H S1 L1 T gh P E
[2021-11-15 16:13:15] one_line_per_rec.cm_anno.srt created
[2021-11-15 16:13:15] Step 6 completed in 0h27m05s

[2021-11-15 16:13:15] Step 7 -- CR_analysis
[2021-11-15 16:13:16] grep -e trnE -e trnF -e ^Glu -e ^Phe hfmt_111521/blast_results/mito_feature_match_to_cand_recs.tsv
[2021-11-15 16:13:16] trnE_trnF_distances.tsv created
[2021-11-15 16:13:17] ControlRegion_btw_trnE_trnF_length.stats created
[2021-11-15 16:13:18] CR1 has type remnant, with mean length 178 bp found from 113 reads containing it. flanking trnas E and F

ControlRegion_btw_trnE_trnF_length.stats
Num CR_len
113 178

recs: 113
mean: 178
stddev: 0
mode: 178
longest: 178
shortest: 178
diff: 0
num within one stddev: 113 100.00%

[2021-11-15 16:13:18] grep -e trnT -e trnP -e ^Thr -e ^Pro hfmt_111521/blast_results/mito_feature_match_to_cand_recs.tsv
[2021-11-15 16:13:18] trnT_trnP_distances.tsv created
[2021-11-15 16:13:19] ControlRegion_btw_trnT_trnP_length.stats created
[2021-11-15 16:13:20] CR2 has type Control Region, with mean length 1430 bp found from 99 reads containing it. flanking trnas T and P

ControlRegion_btw_trnT_trnP_length.stats
Num CR_len
99 1430

recs: 99
mean: 1430
stddev: 0
mode: 1430
longest: 1430
shortest: 1430
diff: 0
num within one stddev: 99 100.00%

[2021-11-15 16:13:20] Step 7 completed in 0h0m05s

[2021-11-15 16:13:20] Step 8 -- split_recs_into_sets
[2021-11-15 16:13:20] created split_sequences to hold sequences for assembly by alignment
[2021-11-15 16:13:21] tRNAs T P flank the Control Region.
[2021-11-15 16:13:21] Splitting sequences from Pro to sequence end.
[2021-11-15 16:13:21] Splitting sequences from sequence beginning to Thr.
[2021-11-15 16:13:21] Splitting sequences from Thr to Pro to capture the Control Region and its flanks
[2021-11-15 16:13:21] Split sequences created for alignment assembly and for Control Region assembly & repeat analysis

[2021-11-15 16:13:21] Step 8 completed in 0h0m01s

[2021-11-15 16:13:21] Step 9 -- assemble_mito
[2021-11-15 16:13:22] Step 9a -- assemble using megahit
[2021-11-15 16:13:22] running megahit to assemble mito records using sequences in mito_hifi_recs.fasta
[2021-11-15 16:13:24] megahit_best.fa mito file created
[2021-11-15 16:13:24] running MiTFi to determine location of Phe so that the sequence will start there
[2021-11-15 16:17:06] MiTFi run on megahit_best.fa completed
[2021-11-15 16:17:06] mito_megahit.fasta reoriented to start with Phe.
[2021-11-15 16:17:06] Running MiTFi on the reoriented sequence.
[2021-11-15 16:20:49] input files checked
[2021-11-15 16:20:49] search for 12S_s-rna
[2021-11-15 16:20:50] search for 16S_l-rna
[2021-11-15 16:20:52] mito_megahit_rrnS.tbl mito_megahit_rrnL.tbl created in megahit_out for mito_megahit.fasta
[2021-11-15 16:20:52] goose hairpin sequence information added, if any found
[2021-11-15 16:20:52] mito_megahit.cm_anno with rrns, rrnL, cr and any goose hairpins added to mitfi results created for mito_megahit.fasta
[2021-11-15 16:20:52] MiTFi annotation of mito_megahit.fasta completed, results in mito_megahit.cm_anno

[2021-11-15 16:20:52] Step 9b -- assemble using multi-sequence alignment (msa) consensus
[2021-11-15 16:20:52] running mafft --auto Pro_to_end.fasta >Pro_to_end.mafft.fa 2>Pro_to_end.mafft.log
[2021-11-15 16:22:11] consensus_from_fasta_alignment.sh Pro_to_end.mafft.fa >Pro_to_end.consensus.fa
[2021-11-15 16:22:11] running multiple sequence alignment on reversed sequences of beg_to_Thr.fasta
[2021-11-15 16:22:11] running mafft --auto reversed_beg_to_Thr.fa >reversed_beg_to_Thr.mafft.fa 2>reversed_beg_to_Thr.mafft.log
[2021-11-15 16:23:32] consensus_from_fasta_alignment.sh reversed_beg_to_Thr.mafft.fa >reversed_beg_to_Thr.consensus.fa
[2021-11-15 16:23:32] re-reversing consensus file to make forward version beg_to_Thr.consensus.fa
[2021-11-15 16:23:32] running mafft --auto Thr_CR_Pro.fasta >Thr_CR_Pro.mafft.fa 2>Thr_CR_Pro.mafft.log
[2021-11-15 16:25:07] consensus_from_fasta_alignment.sh Thr_CR_Pro.mafft.fa >Thr_CR_Pro.consensus.fa
[2021-11-15 16:25:07] non_cr_consensus.fasta from Pro_to_end.consensus.fa & beg_to_Thr.consensus.fa
[2021-11-15 16:25:07] running mafft --localpair <(cat Pro_to_end.consensus.fa beg_to_Thr.consensus.fa) >non_cr_consensus.mafft.fa 2>non_cr_consensus.mafft.log
[2021-11-15 16:25:15] consensus_from_fasta_alignment.sh non_cr_consensus.mafft.fa >non_cr_consensus.fasta
[2021-11-15 16:25:15] running mitfi -code 2 analysis on non_cr_consensus.fasta
[2021-11-15 16:28:57] non_cr_consensus.mitfi created with 22 entries
[2021-11-15 16:28:57] removing 1374 from beginning and removing 0 from the end to create 15721bp mito_msa_no_cr.fasta
[2021-11-15 16:28:57] mito_msa_no_cr.fasta created
[2021-11-15 16:28:57] running mitfi -code 2 analysis on mito_msa_no_cr.fasta
[2021-11-15 16:32:23] mito_msa_no_cr.mitfi created with 22 entries
[2021-11-15 16:32:23] input files checked
[2021-11-15 16:32:23] search for 12S_s-rna
[2021-11-15 16:32:24] search for 16S_l-rna
[2021-11-15 16:32:37] mito_msa_no_cr_rrnS.tbl mito_msa_no_cr_rrnL.tbl created in hfmt_111521/msa_assembly/cm_mitfi for mito_msa_no_cr.fasta
[2021-11-15 16:32:37] mito_msa_no_cr.cm_anno created with rrnS, rrnL, cr and any goose hairpin added to mitfi results
[2021-11-15 16:32:38] running mitfi -code 2 -onlycutoff analysis on Thr_CR_Pro.consensus.fa
[2021-11-15 16:33:01] Thr_CR_Pro.consensus.mitfi created with 2 entries
[2021-11-15 16:33:02] mito_msa.fasta created
[2021-11-15 16:33:02] Control Region followed by Pro (P) but first_trna set to Phe, file will be reflowed to begin with Phe (F)
[2021-11-15 16:33:02] mito_msa.fasta reflowed to begin with Phe (F)
[2021-11-15 16:33:02] running mitfi -code 2 analysis on mito_msa.fasta
[2021-11-15 16:36:44] mito_msa.mitfi created with 22 entries
[2021-11-15 16:36:45] input files checked
[2021-11-15 16:36:45] search for 12S_s-rna
[2021-11-15 16:36:46] search for 16S_l-rna
[2021-11-15 16:36:48] mito_msa_rrnS.tbl mito_msa_rrnL.tbl created in hfmt_111521/msa_assembly/cm_mitfi for mito_msa.fasta
[2021-11-15 16:36:48] mito_msa.cm_anno created with rrnS, rrnL, cr and any goose hairpin added to mitfi results
[2021-11-15 16:36:48] Step 9 completed in 0h23m27s

[2021-11-15 16:36:48] Step 10 -- compare_assemblies
[2021-11-15 16:36:49] compare_megahit_msa created
[2021-11-15 16:36:49] Comparing mito_megahit.fasta (17172 bp) and mito_msa.fasta (17151 bp)

Comparison between mito_megahit.fasta (17172 bp) and mito_msa.fasta (17151 bp)

edit distance 21
17149 bp in runs of matches 100 or greater
15976 matches at the beginning and 1173 at end for 17149 contiguous matches

Following describes how to transform the 17172 bp mito_megahit.fasta into the 17151 bp mito_msa.fasta
15976=
1D1=10D1=10D
1173=

Areas of difference between mito_megahit.fasta and mito_msa.fasta

ACTCTAGGCACTTCCATCTAAAACGTATTAATTCTTGTCATCAATTTTTTTTTCAAAAATTTAAAATTTTTTACACATTCCA TTCCGCCGATTTCCCTCTTTAATAAAATTCATTCTTTTTTTTTGCAAAAATTTTCTTTTTTTTTGCAAAAATTTTTAC mito_megahit.fasta 15841-16000
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||-|----------|----------| 139 matches 21 indels
ACTCTAGGCACTTCCATCTAAAACGTATTAATTCTTGTCATCAATTTTTTTTTCAAAAATTTAAAATTTTTTACACATTCCA TTCCGCCGATTTCCCTCTTTAATAAAATTCATTCTTTTTTTTTGCAAAAATTTT-T----------A----------C mito_msa.fasta 15841-15979

[2021-11-15 16:36:50]	Step 10 completed in 0h0m01s

[2021-11-15 16:36:50]	Step 11 -- assemble_CR
[2021-11-15 16:36:53]	No tandem repeats found in the 99 Control Region sequences and length stddev 0 is less than 20, consensus will be used.
[2021-11-15 16:36:54]	Step 11 completed in 0h0m04s

[2021-11-15 16:36:54]	Step 12 -- complete
[2021-11-15 16:36:54]	msa consensus mitochondrion chosen as best representative
[2021-11-15 16:36:54]	mitochondrion.fasta created
[2021-11-15 16:36:55]	blastn of features against mitochondrion.fasta for protein coding gene annotation
[2021-11-15 16:36:55]	blastn -db hfmt_111521/complete/blastdb/mitochondrion.fasta -query top_match_feature_sequences.fasta -outfmt "6 std staxid stitle qlen qcovhsp qcovus" -task blastn -evalue 1e-10 | sort -k2,2V -k9,9n -k12,12nr
[2021-11-15 16:36:57]	mitochondrion.anno created
[2021-11-15 16:36:57]	Step 12 completed in 0h0m03s

[2021-11-15 16:36:57]	1h8m13s to complete Mon 15 Nov 2021 04:36:57 PM PST

settings.tsv:
program_title HiFiMiTie version 0.01 -- Find & Analyze Metazoan Mitochondria from HiFi reads
version 0.01
version_date 15-Oct-2021
run_by jdumbacher
mitodb_dir /ccg/blastdbs
mitodb_name mito
taxonomy_dir /ccg/db_sets/taxdump
working_dir *workdir*/pyrocephalus/mito/hfmt_111521
start Mon 15 Nov 2021 03:28:44 PM PST
threads 32
taxid 8782
taxname Aves
taxlineage /ccg/db_sets/taxdump/fullnamelineage.dmp
mitogenomes 964
step 12
code 2
step_completed Mon 15 Nov 2021 04:36:57 PM PST
HiFi_mito_reads 167
top_mito_matches NC_007975
top_mito_match_first_recname Phe_NC_007975
top_mito_match_first_trna F
taxname_trna_starts Aves: F 888, V 32, T 20, I 10, S2 10, E 2, G 1, P 1
taxname_first_trna F
first_trna F
last_trna E
mito_blast_last_trna_counts E:113
trna_order F V L2 I Q M W A N C Y S2 D K G R H S1 L1 T P E
cm_anno_order F 12S V 16S L2 I Q M W A N C Y S2 D K G R H S1 L1 T gh P E
gh_prev_trna T
gh_succ_trna P
gh_rec_count 114
gh_found_in_recs 114
num_CRs 2
Primary_CR CR2
CR1_flanks E F
CR1_mean 178
CR1_stddev 0
CR1_recs_w_CR 113
CR1_type remnant
CR2_flanks T P
CR2_mean 1430
CR2_stddev 0
CR2_recs_w_CR 99
CR2_type Control Region
msa_megahit_edit_dist 21
msa_megahit_edit_cigar 15976=1D1=10D1=10D1173=
CR2_seqs_w_repeats 0
assemble_CR_result msa
anno_order F rrnS V rrnL L2 ND1 I Q M ND2 W A N C Y COX1 S2 D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4
H S1 L1 ND5 CYTB T cr gh P ND6 E cr
finished Mon 15 Nov 2021 04:36:57 PM PST
run_time 1h8m13s

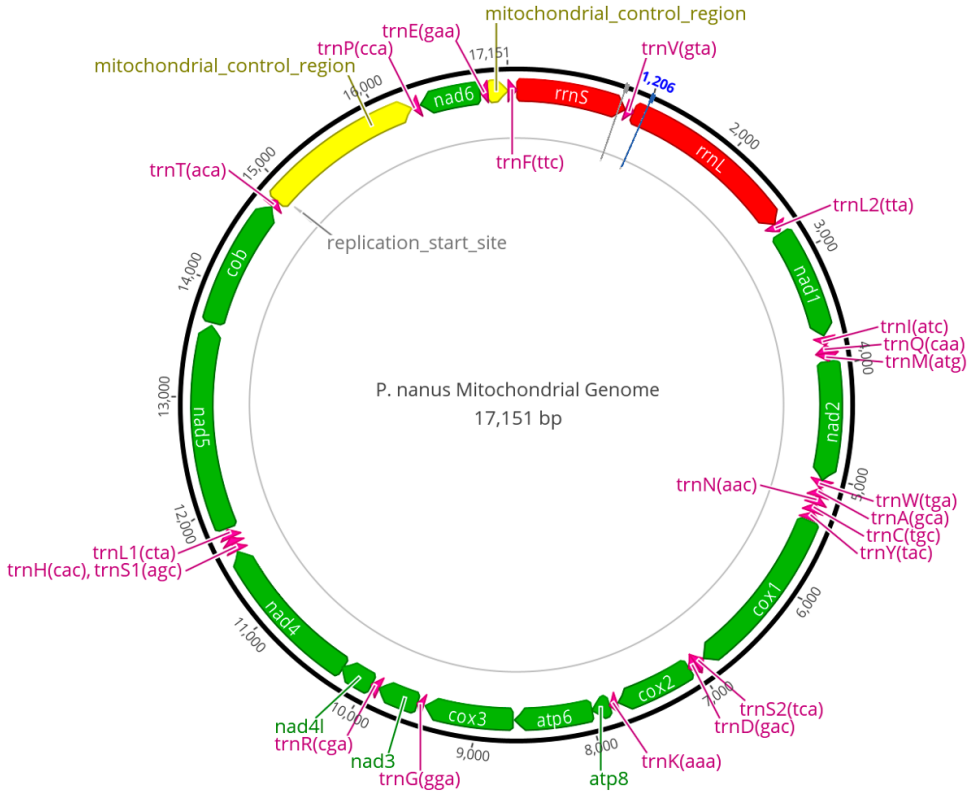


Fig S1. Complete mitochondrial genome was 17151 nucleotides in length as derived from a total of 167 corrected HiFi reads that mapped to the mitogenome with 13 protein coding genes, 22 tRNAs, 2 rRNAs, and 2 control regions, one complete, the other designated as a remnant CR.

Mitochondrial DNA genome (in fasta text format):

>Pnanus_complete_mitochondrion length 17151
GTCCCTGTAGCTTACAAAAAGCATAACACTGAAGATGTTAAGATGGTTGCCCTAAACACCCAAGGACAAAA
GACTTAGTCCTAACCTTACTGTTAGTTCTTGCCACACATATACATGCAA
GTATCCGCACTCCAGTGAAAATGCCCTCGACACCTTAAAAAGATAGTAGGAGCAGGTATCAGGCTCACTTAA
CTGTAGCCCAAAACGCCTTGCCTAGCCACACCCCCACGGGTACTCAGC
AGTAATTGACATTAAGCAATAAGTGTAAACTTGACTTAGTTAAAGCAACCTACTAGGGTTGGTAAATCTTGT
GCCAGCCACCGCGGTCATACAAGAAACCCAAATTAACTTTTCACGGCG
TAAAGAGTGGTCTCTCATTATCACCTCAACTAAGATTGAAATGCAATCAAGCTGTTATAAGCATAGAATGCA
CTTAACATCCCCTATCAAAATGATCTTAGCCCCCCGACTAATAAAGCC
CACGAAAGCCAGGTCACAAACTGGGATTAGATACCCCATTATGCCTCGCCCTAAATCCTGATGTTTCTCCTAC
CAAAACATCCGCCCGAGAACTACGAGCACAAACGCTTAAAACTCTAA
GGACTTGGCGGTGCTTCAAACCCACCTAGAGGAGCCTGTTCTATAATCGATAACCCACGATATACCCAACCA
CTTCTTGCCAAATCAGCCTATATACCGCCGTCGCCAGCTCACCCCTAC
CCTGAGGGCCTAACAGTGAGCATAATAGCCCCCCCGCTAGCAAGACAGGTCAAGGTATAGCTTATGAAGTG
GAAGAAATGGGCTACATTTTCTAATATAGAAAACCACCCTACGACAAGG
AACATGAAACTATTCCTAAAAGGCGGATTTAGTAGTAAAGCAGGACAATCATGCCTTCTTTAAACCGGCTCT
GGAGCACGTACATACCGCCCGTCACCCTCCTCACAAGCTACTTCACTA
TATTCATACATAATTCACTTTTAAGCTAAAGATGAGGCAAGTCGTAACAAGGTAAGTGTACCGGAAGGTGTA
CTTAGTCTACCAAGATGTAGCTATAACAAAAGCATTCAGCTTACACCT
GAAAGATACCTGTCCAGTATCAGGTCATCTTGAAGCCTACTCTAGCCCGCTCTTCCATCCACATAAACCAATT
GCCCTCTCTTCAAAAACCAAAACATTTTTTTACCTCAGTATAGGCGA
TAGAAAGGTACCCCAAGGAGCAATAGAGACCACGTACCGCAAGGGAAAGATGAAATAGTAATGAAAACCC
AAGCGCCAAATAGCAAAGTTCAATCCTTGTACCTTTTGCATCATGATTTA
GCAAGAACTGACCAAGCAAAACGCTAATTTAAGCTTGCCACCCCGAAACCCGCGCGAGCTACTTACGAGCA
GCTATTCAATGAGCAAACCCGTCTCTGTTGCAAAAGAGTGGGACGACTC

GTCAGTAGAGGTGAAAAGCCAATCGAGCCGGGCGATAGCTGGTTGCCTACGAAACGAATCTTAGTTCACCCT
CAATCTCCCTCCAAGGCCCAAACCAAACCCCCATGAAGAAGATCGAGG
GCTACTTAAAGGAGGTACAGCTCCTTTAAAAAGAACACACTCTCCACAAGCGGATAAGTTCTCCCATATCCT
AACTGTGGGCCTTAGAGCAGCCACCACCAAAGAATGCGTCAAAGCTCA
ACCCCTAAAAATTTGAAAGCCCTACAACTCCCTCTCCCCTAATAGGCTAACTTATATCTATAAGAGAATCAAT
GCTAAAATGAGTAACTAGGAACCTCCTCTATGGCGTAAACTTACATC
CCCACATTATTACCAAATTTAGAGGTATATACTACCCTAACAAGACCCAATACCATAAAACCTGTTAACCCA
CCCAAGGAACGCCCCAAAGAAAGATTAAAATCTGCAAAAGGAACTAGG
CAACTCCAAGGCCCGACTGTTTACCAAAAACATAGCCTTCAGCGAGTCAAGTATTGAAGGTGAAGCCTGCCC
AGTGACAGTACGTTCAACGGCCGCGGTATCCTAACCGTGCGAAGGTAG
CGCAATCAATTGTCCCATAAATCGAGACACGTATGAATGGCTAAACGAGGTCTTAACTGTCTCTTGCAGATA
ATCAGTGAAATTGATCTCCCTGTGCAAAAGCAGGGATCATAACATAAG
ACAAGAAGACCCTGTGGAACTTCAAAATCAACAGCCACTCCCATATACCACTCCCCCTACATATGGAATACC
TCACGGGACTCTATTGGCCTGTATTTTTTGGTTGGGGCGACCTTGGAG
AAAAACGAATCCTCCAAAAATTAGACCACCCCTCTACACCAAGAACAACCTATCAACGTACTAACAGTAACC
AGACCCAATACCATTGATTAATGGACCTAGCTACCCCAGGGATAACAG
CGCAATCTCCCCCAAGAGCCCCAATCGACAGGGAGGTTTACGACCTCGATGTTGGATCAGGACATCCTAGTG
GTGCAGCCGCTACTAAGGGTTCGTTTGTTCAACGATTAACAGTCCTAC
GTGATCTGAGTTCAGACCGGAGTAATCCAGGTCGGTTTCTATCTATGATAAACTCCTTCCAGTACGAAAGGA
CAGAAAGAGTAAGGCCAATACCCCAAGCACGCCTTCGCCCAAAGATAT
GAACTCAACTAAATCTCTTAGGGCTAACTTTCCTTACCCTCCTAAAAAAGGACCGCTAGAGTGGCAGAACTT
GGCAAATGCAAAAGGCTTAAGCCCTTTATCTAGAGGTTCAAATCCCCT
CTCTAGCTTCCACACCAACTATATGATAATTTACCCCAACCTAACTCATCTCATCATATCCTTATCCTATGCCT
TACCCATTCTAGTTGCCGTAGCCTTCTTAACCCTAGTAGAACGTAA
AGTCCTAAGCTATATACAAACACGTAAAGGCCCAAACATCGTAGGACCATTTGGTCTTCTACAACCTATAGC
AGATGGTGTAAAACTTTTTATTAAAGAGCCAATCCGCCCATCCACTTC
TTCCCCCTACCTATTTATTCTCACTCCAATATTGGCCCTCCTTCTAGCAGTCATAATCTGAACCCCACTCCCAC
TACCTTTTCCTCTTGCAGATATAAACCTAGGCATACTTTTCCTCCT
CGCACTATCTAGCCTCGCAGTCTATTCAATTCTATGATCTGGATGGGCCTCAAACTCAAAATATGCACTAATC
GGTGCTCTACGAGCAGTCGCACAAACAATTTCCTACGAAGTCACATT
AGCCATTATTCTCCTATGTATTATCATTTTTAGCGGAAACTATACCCTTAATACCCTTACTACAACCCAAGAA
CCAATATACCTCATATTCTCTTCTTGACCCCTAGCAATAATATGATA
TATTTCCACCCTCGCAGAAACAAACCGTGCCCCATTTGATCTAACAGAGGGTGAATCAGAACTAGTCTCAGG
GTTTAACGTTGAATATGCTGCAGGACCCTTTGCCTTATTTTTCCTAGC
TGAATACGCTAACATCATACTTATAAATGCACTAACTACTATCTTATTCTTAAACCCAAGTTCACTTAACATC
ACCCAAGAACTCTACCCACTAACTCTAGCTACTAAAACCCTCCTACT
TTCCGCAGGATTCCTATGAATCCGCGCCTCCTACCCTCGATTCCGCTACGACCAACTCATGCACCTACTATGA
AAAAGTTTCCTGCCACTTACACTATCATTATGCCTATGACATATCAG
CCTTCCAATCTCATTCGCAGGTCTACCACCTCACCTAAGACTCCCGAGGAAATGTGCCTGAACGTTAAAGGG
TCACTATGATAAGGTGAACATAGAGGTATACCAACCCTCTCATTTCCT
AGACTTAGAAAAGTAGGAATCGAACCTACACAGAAGGAATCAAAATCCTTCATACTTCCTTTATATTATTTC
CTAGTAAGGTCAGCTAATCAAGCTATCGGGCCCATACCCCGAAAATGA
TGGTTTAACTCCTTCCCTTACTAATAAACCCCCAAGCTAAATTCATTTTCTCAATAAGTCTCCTCCTAGGTACT
ACAATTACAATTTCAAGCAACCACTGAATTATAGCATGAGCTGGAC
TTGAAATTAATACTTTAGCTATCCTCCCCCTAATCTCAAAATCTCATCACCCACGAGCCATTGAAGCCTCAAC
TAAATACTTCCTAGTCCAAGCAACCGCCTCCACACTCCTCCTATTTT
CCAGTATAACCAACGCATGATTCACTGGTCAGTGAGATATTACTCAACTCACTCATTCAGTATCATGTCTACT
GCTAACAGCTGCAATCTCAATAAAGCTAGGCTTGGTCCCATTCCACT
TCTGATTCCCAGAAGTCCTTCAAGGTTCCTCCTTAATAACTAGCCTGCTGCTAGCCACAATCATAAAATTCCC
TCCCACCGTACTCCTCTTGTTAACCTCCCCCTCACTAAATCCCACAC
TACTATCCATATTAGCAATTGCCTCCGCTGCCCTAGGAGGCTGAATAGGGCTCAACCAAACTCAAATCCGCA
AAATTATAGCCTTCTCGTCTATCTCCCATCTAGGCTGAATGACCATTA
TCCTCATCTACAACCCTAAACTCACGCTTATTGCTTTCTACCTCTATTCTCTAACTACAGCTGCCATTTTCTCT
GCCCTCAGTGCTATCAACTCTTTAAAACTGACCACCTTAATAACTG
CGTGATCCAAAATTCCTGCACTAAGTGCAACCCTAATACTAACTCTTCTATCTCTCGCAGGTCTTCCCCCATT
AACTGGATTTCTACCAAAATGGCTAATTATTCAAGAGCTAACTAAAC
AAGAACTAACAGCCACAGCTACTATCATTGCCCTACTCTCTCTCCTAGGACTTTTCTTCTACCTCCGTCTTGCT
TACTGTGCAACAATTACACTTCCCCCAAACTCTGCTAACCACATAA
AACAATGACGAACCAACAAGCTCACTAACTTACTAACCCCCACCCTCATTATAATATCAGCTACACTTTTACC
CCTATCACCTACAATCCTTACTATTCCATAGAAGCTTAGGTTACTTA
AACCGAAGGCCTTCAAAGCCTTAGACAAGAGCTATTTCCTCTTAGCTTCTGCTAAGATCCGCAGGATACTAA
CCTGCATCCTCTGAATGCAACCCAGATGCTTTAACTAAGCTAGGACCT

TCCCTAGACAGATGGGCTTCGATCCCATAACACTCTAATTAACAGCTAGATGCCATAACCTACAGGCTCCTG
TCTACTAGACTCCGGTATATACTTAATACACATCTTTGAGTTTGCAAC
TCAATATGAAATTCACTACAGAGTCGATAAGAAGAGGAATTAAACCTCTGTCAAAAGGACTACAGCCTAAC
GCTTCAACACTCAGCCATCTTACCTATGACCTTCATTAACCGATGATTA
TTCTCAACCAACCATAAAGACATTGGCACATTATACCTAATTTTTGGCGCTTGAGCTGGCATAATTGGTACCG
CCCTAAGCCTCCTTATCCGAGCAGAATTAGGACAACCAGGGACTCTC
CTAGGGGATGACCAAATCTATAATGTAATCGTCACTGCTCATGCCTTTGTAATAATCTTCTTTATAGTAATGC
CCATTATAATCGGAGGATTCGGTAATTGATTAGTTCCCTTAATAATT
GGTGCCCCCGACATAGCATTCCCCCGCATGAACAACATAAGTTTCTGACTATTGCCCCCATCATTCCTCCTTC
TTCTAGCCTCATCCACAGTCGAAGCCGGAGCCGGAACAGGATGAACT
GTCTACCCACCATTAGCTGGAAACCTAGCACATGCTGGAGCTTCCGTAGACTTAGCTATTTTCTCCCTTCACC
TTGCAGGTGTCTCTTCAATTTTAGGTGCCATCAATTTTATTACTACC
GCAATTAACATAAAACCACCCGCCCTATCACAATACCAAACTCCCCTCTTTGTGTGATCCGTCCTAATCACTG
CAGTCCTTCTCCTCCTCTCTTTACCAGTCCTTGCCGCAGGTATCACC
ATGCTATTAACAGATCGCAACCTTAACACTACATTTTTCGACCCCGCAGGAGGAGGAGACCCAATCTTATAC
CAACATCTTTTCTGATTCTTTGGTCACCCCGAAGTTTATATTCTAATC
TTACCAGGATTCGGTATTATTTCTCATGTAGTAGCATACTACGCTGGCAAAAAAGAACCATTCGGCTACATA
GGAATAGTCTGAGCCATACTCTCTATTGGTTTCCTAGGTTTTATCGTA
TGAGCTCATCATATATTTACAGTAGGAATGGACGTAGACACTCGAGCATACTTTACATCCGCCACAATAATC
ATTGCAATCCCTACCGGCATTAAAGTTTTTAGCTGACTAGCTACACTT
CATGGAGGGACTATTAAATGAGACCCTCCCATGCTATGAGCTCTGGGATTTATCTTTCTATTTACCATCGGAG
GACTCACAGGCATTGTCTTAGCCAACTCTTCCTTAGACATTGCTCTT
CATGACACTTATTATGTAGTAGCCCACTTCCACTATGTCTTATCAATAGGGGCTGTATTTGCAATCCTAGCAG
GCTTTACACACTGATTCCCACTCTTCACAGGCTACACCCTCCATCAA
ACATGAGCTAAAGCCCATTTTGGAGTCATGTTTACAGGTGTAAATTTAACCTTCTTCCCCCAACACTTCCTAG
GCCTAGCTGGTATACCACGACGATATTCAGATTACCCAGATGCCTAC
ACCTTATGAAACACCCTATCCTCTATTGGCTCTCTCATCTCTATAACTGCTGTAATTATGCTTATGTTTATCAT
CTGAGAAGCTTTTGCATCCAAACGAAAAGTAATGCAACCTGAACTT
ACTTCTACCAACATTGAATGAATCCATGGTTGCCCTCCCCCATATCACACTTTTGAAGAACCAGCCTTCGTCC
AAATCCAAGAAAGGAAGGAATCGAACCCTCACAAGCTGGTTTCAAGC
CAACCGCATTAAGCCACTTATGCTTCTTTCTTATGGGTTGTTAGTAAAAAAATTACATAGTCTTGTCAAGACT
AAATCACAGGTGAAAACCCAGTACACCCCATCCCTTCCAATGGCCAA
CCACTCCCAATTCGGTTTCCAAGACGCCTCATCTCCCATCATAGAAGAATTAATTGAATTCCATGATCACGCT
TTAATAGTTGCCCTAGCAATCTGTAGCCTAGTACTTTACCTCCTGAC
CCTTATACTCATAGAAAAACTATCATCCAACACTGTTGACGCGCAAGAAGTAGAACTAGTCTGAACCATTCT
ACCCGCAATTGTCCTAATCATGCTAGCCCTTCCCTCACTTCAAATTCT
TTATATAATAGACGAAATTGACGAACCCGACCTCACACTTAAAGCTATCGGCCACCAATGATATTGATCCTA
TGAATATACAGACTTCAAAGAATTATCCTTCGATTCATACATACTACC
AACTACAGAACTCCCACTAGGTCACTTTCGACTACTAGAAGTTGACCACCGTATTGTTATTCCTATAGAATCC
CCCATTCGGGTTATCGTAACAGCCGATGACGTCCTACATTCTTGAGC
CATTCCTAGTCTCGGCGTAAAAACTGACGCAATCCCAGGGCGACTCAACCAAACATCATTTATTACTACCCG
ACCTGGAATTTTCTATGGTCAATGTTCAGAAATCTGCGGAGCGAACCA
TAGCTTCATACCTATCGTAGTTGAATCAGCCCCCCTAACCCACTTCGAAAGCTGATCTTCACTCCTATCATCA
TAAACATTAAGAAGCTATGAACCAGCACTAGCCTTTTAAGCTAGAGA
AAGAGGACCTCAACCCTCCTTAATGGTATGCCACAGCTAAACCCAGCACCATGATTCTTTATCATGTTATCCT
CATGATTTATTTTTTCCTTTATCCTTCAGCCCAAAATTCTATCATTC
ACCTCCACAAACCAACCCTCTAACAAAGCCCTCTCTACCATAAAAACATCCCCCTGAAACTGACCATGAACC
TAAGCTTCTTTGACCAATTTATAAGCCCTTCCCTATTAGGAATCCCAT
TAATCCTTATCTCACTCTTGTTTCCAACACTCCTATACCCTTCCCCTAGCAACCGATGAATTACAAACCGTATT
TCTACCCTCCAACTCTGACTCTTTCAACTACTCACAAAACAACTTA
TAACCCCAATAAACAAAAAAGGACATAAATGAGCCCTAATCTTACTCTCACTCATAATCTTTCTCCTGACAA
TTAATCTTCTAGGGCTCCTACCTTATACATTCACTCCTACTACCCAAT
TATCCATAAACCTAGCCCTAGCCTTTCCACTATGACTCGCAACCTTACTCACAGGCCTACGAAACCAACCCTC
AGCCACTCTAGGCCATCTCCTGCCAGAAGGCACCCCCACACCCTTAA
TTCCTGCCCTCATTTTAATTGAAACTACTAGCTTACTTATTCGCCCCTTAGCCCTAGGAGTTCGCCTTACTGCT
AACCTCACAGCCGGACATTTACTCATCCAACTTATTTCAACTGCTA
CTACTGTCCTTCTACCTATAATTCCAACAGTATCTTTACTTACTGCAACCATCCTCTTTTTACTCACTATCTTA
GAGGTAGCTGTTGCCATAATTCAAGCCTACGTTTTCGTTCTACTAT
TAAGTCTTTACCTACAAGAAAATATCTAAACCCTAATGACCCACCAAGCACACTCATTTCATATAGTAGATC
CAAGCCCCTGACCCATTTTTGGAGCAGCAGCTGCTCTTCTCACTACTT
CAGGACTAACCATATGATTTCATCATAACTCTACCCAACTCCTAACCTTAGGGTTATTATCCATAGTCCTTGT
TATAATTCAATGATGGCGCGATATCGTACGAGAAAGCACATTCCAAG

GCCATCATACTCCTACCGTCCAAAAAGGTCTACGTTACGGTATAATCTTATTTATTACATCCGAAGCATTTTT
TTTCTTAGGCTTCTTCTGGGCATTCTTCCATTCAAGCCTAGTCCCAA
CTCCAGAACTCGGAGGACAATGACCTCCTATAGGAATCAAACCCCTCAACCCCATAGAAGTTCCCCTATTAA
ATACAGCGATCCTACTAGCCTCTGGTATTACCGTCACATGAGCACACC
ATAGCATTGTAGAAGCCAACCGAACCCAAGCAACCCAAGCTCTATTTATAACTATCGCGTTAGGATTCTACT
TCACAGCTCTCCAAGCAATAGAATATTATGAAGCCCCATTTTCAATCG
CCGATGGAGTCTACGGATCAACTTTTTTTGTAGCCACAGGATTTCATGGCCTTCATGTAATCATTGGTTCTTC
CTTCCTATCCGTCTGCTTCCTTCGACTTGTTAAATTCCACTTCACCT
CTAACCACCACTTTGGTTTCGAAGCAGCAGCTTGATACTGACATTTTGTTGATGTCATCTGACTATTCCTCTAT
ATAACTATTTACTGATGAGGATCCTGCTCTTCTAGTATAATAATTA
CAATTGACTTCCAATCTTTAAAATCTGGACCAAAACCAGAGAAGAGCAATTAACATAATCACCTTCATACTG
ATATTATCCCTATCACTAAGTATCATCCTCACTACACTAAATCTATGA
CTAGCACAAACCAACCCAGACTCAGAAAAATTATCCCCCTACGAATGCGGCTTCGACCCACTAGGATCCGCT
CGACTCCCATTTTCAATCCGATTCTTTCTAGTCGCAATTTTATTCCTC
CTCTTTGACTTAGAAATTGCACTCCTTTTACCCCTCCCGTGAGCTACTCAACTTCAATCCCCCCTCACCACCCT
AACATGGGCATTTATCATACTTCTTCTCCTAACACTAGGACTTATA
TATGAATGAGCCCAAGGAGGTCTAGAATGGGCAGAATAACTCCAGAAAGTTAGTCTAATAAAGACAGTTGA
TTTCGGCTCAACAAATCACAGTCCTACCCTGTGACCTTCTCTATGACCT
TCCTACACCTAACTTTCTATTCCACTTTTACCCTAAGCAGCCTAGGCCTAGCCTTCCACCGAACCCACCTAAT
CTCTGCCCTTTTATGCTTAGAAAGCATAATACTGTCTATATATATCG
CCCTTTCTTTATGACCCATCCAAACACAAATATCCTCCTCCACCCTAATACCCATCCTAATATTAGCATTCTCT
GCCTGCGAAGCAGGCGCTGGCCTTGCAATTCTAGTAGCCTCAACCC
GGACCCATGGATCCGACCACCTCCACAACCTCAACCTCTTACAATGCTAAAAATTATCATCCCAACGATAAT
ACTCCTCCCAACAGCCCTCCTATCACCCTCAAAGTCCCTATGAATCAA
TGTTACCTCCCACAGCCTATTAATTGCAGCCGTAAGCCTCCATTGACTCTCCCCAACATACTACGCTAACAAA
AATCTATCTCAATGACTTGGGGTGGATCAAATCTCATCACCATTACT
AGTCTTATCTTGCTGACTTCTACCCCTCATAATATTAGCTAGCCAAAACCACCTTCAACAAGAACCATCCACA
CGAAAACATATTTTTATCCTAACCATAATTCTTACTCAACCCTTTAT
TTTACTAGCATTCTCATCCCTAGAACTCATATTATTTTATATTGCATTCGAAGCTACCTTAATCCCCACCCTCA
TCCTAGTTACACGGTGAGGAAGTCAACCAGAACGCCTAAGTGCTGG
TATTTACCTACTATTTTACACACTTATTAGCTCCCTCCCACTATTAGTTGTTATCCTCCACCTTCATACCCAAA
CGGGCACCATGCATCTTGCTATAATAAAACTTATCCACCCGTCCCT
TACTGACTCATGAACAGGAACTTTATCCAGTCTCGCCTTACTCCTTGCATTCATAGTAAAAGCTCCTCTTTAC
GGACTACACTTATGACTTCCAAAAGCACACGTAGAAGCCCCAATCGC
AGGATCTATGCTTTTAGCCGCCCTCCTTCTAAAACTTGGAGGATATGGCATCATACGCGTATCCATAATTACA
AACCCCCTCCTAAACTACCTACACTACCCATTTCTATCTCTAGCTCT
ATGAGGAGCATTAATAACAAGCTCCATCTGTCTCCGCCAAATCGACTTAAAATCTATAATCGCATATTCTTCT
GTAAGCCACATAGGCCTAGTCGTAGCCGCAGTCATAATCCAAACCCA
CTGATCATTTTCAGGGGCTATAATCCTAATAATCTCCCATGGACTTACATCCTCCATATTATTCTGTCTAGCCA
ACACAAACTATGAACGAACACATAGCCGAATTCTATTCCTAACACG
AGGCCTACAACCCCTCCTTCCCCTCATAGCTATCTGATGACTTCTAGCCAGTCTCACTAACATAGCACTTCCC
CCCACAACCAACCTTATAGCCGAACTAACCATTATAATTGCCCTCTT
TAACTGATCTGCTCTCACCATTATCTTAACCGGAATTGCAACCCTCTTAACTGCCCTATACACCCTATACATG
TTACTTACAACCCAGCGAGGCCCCATCCCAACTTACATGACATCAAT
TCAAAACTCCAACACCCGAGAACACTTACTTATAACTCTCCACATCCTCCCCGCCCTTTTACTTATTCTAAAA
CCAGAACTTATCTCAATAATCCCTTCATGCAAGCATAGTTTTAATCC
AAACATTAGACTGTGATCCTAAAAATAGAAGTTAAACTCTTCTTGCCTGCCGAGGGGAGGTTTAACCAACAA
GAACTGCTAACTCTTGCATCTGAGTATAAAAGCTCAGCCCCCTTACTT
TTAAAGGATAGTAGCAATCCACTGGTCTTAGGAGCCATTTACCTTGGTGCAAATCCAAGTAAAAGTAGTGCA
TATCCCATTAATCCTAAACTCCTTCATAATCCTCACCCTTCTCATCCT
ACTAACCCCAATCTTCCTTCCCCTTCTAATCAAAAACTACCAAAACTCTCCAACTAACATTACACAAACTGTC
AAAGCCTCATTTATTACCAGCCTAGTGCCAATAACTCTTTTCATATA
TATAAACATGGAAAGTATTTCATTATACTGAGAATGAAAATTTATTATAAACTTCAAAATCCCACTAAGCTTC
AAAATAGACCAATACTCACTAACATTCTTTCCCATCGCATTATTTGT
AACATGATCTATTCTTCAATTTGCATCATGATACATAGCCTCAGAACCCTTCATCACAAAATTCTTCTTCTATC
TTCTAACCTTTCTTATCGCTATACTTTGCCTCATCATCGCAAACAA
CATATTCCTTTTATTTATTGGATGAGAAGGAGTCGGAATCATATCTTTTCTCCTTATCGGATGATGACACGGA
CGAGCAGAGGCCAACACAGCTGCTCTCCAAGCCGTACTCTATAACCG
CATCGGAGACATTGGCCTAATCCTATCAATGGCCTGACTAGCTTCCTCCCTAAACACCTGAGAAATTCAACA
ACTGTCATTTCAGAATCAAATTCCACTACTCCCCCTTCTAGGCCTAAT
CCTCGCAGCCACAGGAAAATCCGCTCAATTCGGTCTACATCCATGACTCCCAGCAGCCATAGAAGGCCCCAC
TCCTGTCTCCGCCCTACTTCATTCTAGCACTATAGTAGTAGCTGGAAT

CTTCTTACTTATCCGCACCCACCCCATACTTTGCAACAACCAAATTGCCCTTACTTTGTGTCTCTGCCTGGGAG
CTCTCTCCACACTATTTGCCGCCGCATGCGCTCTTACCCAAAATGA
CATCAAAAAAATCATCGCCTTCTCCACATCAAGCCAATTAGGCCTAATGATAGTAACAATTGGCCTAAACCT
TCCACAATTAGCCTTCCTACACATTTCAACTCATGCATTCTTCAAAGC
AATACTCTTCTTATGCTCAGGCTCAATCATTCATAGCCTCAACGGTGAACAAGACATTCGAAAAATAGGAGG
CCTACAAAAATCACTTCCAATCACCATATCATGCTTAACCATTGGCAA
CCTAGCCCTCATAGGAACCCCCTTCTTAGCAGGATTCTACTCAAAAGACCTTATCATCGAAAACCTAAACAC
ATCTTACCTCAACACGTGAGCACTTCTCTTAACACTCTTAGCCACATC
CTTTACTGCAACTTACAGCATACGCATATCCTTACTAGTACAAACCAACTTCACCCGAACTTCCTCCTCCACT
CCAATCAATGAAAACCACCCAGCAGTCATAAACCCAATTAGCCGCCT
TGCCATTGGAAGCATCATAGCCGGACTACTAATCACCTCCTTTATCCTTCCTACAAAAACTCCACCTATAACT
ATATCTTTTACTACAAAAACTGCAGCCATCATTGTATCCCTACTAGG
AATCATAATTGCACTAGAACTATCAAAACTAACCCATATTCTCACTAGCCCTAAACAAAATGCCACTACTAA
CTTCTCTACCTCCTTAGGCTACTTCAACCCCCTAATACACCGCCTAAG
TTCAACCAAACTACTAAGCACCGGACAAAATATCGCATCACACCTAATCGACCTGTACTGGTATAAAAAAAT
AGGCCCCGAAGGACTCGCAACCCTCCAACTTAAAGCTACCAAAACCTC
AACTACTCTCCACAGCGGATTAATCAAAACTTACCTAGGATCCTTCGCACTATCTATTTTTATCATTATCCTGT
ACATACACGGAACTAATCAATAATGGCACCCAACCTACGAAAACAC
CACCCTCTCCTAAAGATAGTAAACGACTCTCTTATCGATCTTCCCACTCCATCAAACATCTCAGCCTGATGAA
ACTTTGGTTCTCTTCTAGGAATCTGCCTAGCAACACAAATCGTCACC
GGCCTCCTACTTGCTATACACTACACTGCAGACACCTCTTTAGCCTTTACATCAGTTGCCCACACATGCCGAA
ACGTCCAATTTGGCTGACTAATCCGTAACCTCCATGCAAATGGAGCA
TCCTTTTTCTTCATCTGCATTTACCTACATATCGGGCGAGGATTCTACTATGGATCATACTTATATAAAGAAA
CCTGAAACACTGGCGTTATCTTACTTCTAACCCTAATAGCAACTGCT
TTCGTCGGATATGTCCTTCCATGAGGGCAAATATCATTCTGAGGCGCTACAGTAATTACCAATTTATTTTCAG
CAATCCCTTACATCGGTCAAACACTCGTAGAATGAGCCTGAGGTGGA
TTTTCAGTTGATAACCCTACACTTACTCGATTCTTTGCCCTCCACTTCCTTCTCCCATTCATAATCGCAGGTCT
TACATTTATCCACCTAACCTTCCTTCATGAAACAGGATCAAACAAC
CCTTTAGGAATTTCCTCAAACTGCGATAAAATTCCATTCCACCCTTATTTTTCCACAAAAGATATCTTAGGCT
TCATTATTCTACTCTTACCACTAATAACATTAGCTATATTCTCACCT
AATCTCCTAGGCGACCCAGAAAATTTTACGCCCGCCAACCCATTAGTAACTCCCCCCCACATTAAACCAGAA
TGATATTTCCTTTTTGCGTATGCTATCCTTCGATCCATCCCTAACAAG
CTCGGAGGGGTCTTAGCACTTGCAGCCTCAGTCCTAGTACTATTTCTAGTCCCATTCCTTCACATATCGAAAC
AACGCACCATAACCTTCCGCCCTCTCTCCCAACTACTATTCTGAATC
CTAGTAACTAACCTCCTCATCCTAACATGAATTGGAAGCCAACCAGTAGAACATCCATTCATTATTATCGGCC
AACTGGCCTCACTAACTTACTTCACTATCATTCTAATCCTATTCCCA
ATTATCGAAATACTAGAAAACAAGCTACTAAGTTTCTAACTACTCTAATAGTTTATTAAAACATTGGTCTTGT
AAACCAAAGACTGAAGGTTCCCCCCTTCTTAGAGTTATAAACCCCTA
CAACTTAACACCAGCACACACTTCCCTCTAATAATAAAAGGGACCCCCCCCCTCCCCCCCAATTTTTACATTT
TAGGGTATGTATTACTTTGCATTACATTATTTTCCACATTAGACATA
CCATGCATGTAGGAAAATGTCATATTAATTTAATGGTCGGAGCGCATAAATTTTCATGCTTAGTCCCATAACA
ATCCACCCAAGCCATATCCCGATCTAGGCACATTTCTACTTCAAGGA
CCCGCCATGTCATGATCTAGGAATATTCCCAATACCCGGACTAAAACCTATTAAATGCCAGTTTTTGCATAAA
TCCTTCTTATACACGAGGAACCTCCCAAGACTTAAAATCCATGTACT
AAGGACCACTCATCAATTAACCTTGCTCTACGTACTGACCAAGAAGTACTGGGTTATTTATTGATCGTACCCC
TCACGAGAAACCAGCAACCCGGCGTTAGTAATGTTTATCACGACCAG
CTTCAGGTGCATTCTTCCCTCGTCCCTGGCCCAACTTGCGCTTTTGCGCCTCTGGTTCCTCGGTCAGGGCCATA
ACTTGATTAATTTTCCTATACTTGCTCTCCGTTACTAATGGTTGGG
GATGCTTGACCATTATGGACCTCGTGATCGCGGCATCTTCTTTCTTCTATACTTCTCTTATCTTTTTTGGGGTA
GATCTCAATAAGCCCTTCAGAGTGCTCCGCCAGTGGGCCCCCATAT
ATGCTTGACATGTCCATCTTGTGGTCGGCGTGCGGTTTTCTCACTTTCTCGAGCAAATTAATGATATGGGGTC
ATGTGTAACTCGGTCTCATACTGTAGCACTGATGCACTTTGTTGTGC
TATTGGTTTGGAACTTCCAGTTCGTCTCTAAGCTAGGTGTTGTTCAGTTAATGCTCGCCGGACATATTTTGAC
CTCGTCAATTTCACAAAAATTGACAAAAATTTATACATTTTGCCACT
ACTCTAGGCACTTCCATCTAAAACGTATTAATTCTTGTCATCAATTTTTTTTTCAAAAATTTAAAATTTTTTAC
ACATTCCATTCCGCCGATTTCCCTCTTTAATAAAATTCATTCTTTT
TTTTTGCAAAAATTTTTACATTCAATTTAAAAACCCATGAATTTTCTAGGCTAATCCCTTTTATTATCATCGTT
TTATGTTTATTTATTTATTTTTGCATAACCATCCATCAATTTTTCT
ACTAGAATCCCTTTAATCAACGAATTATTCAACTTACCCCCACAAAATTTCCAACGATATATTTATTGTTTTG
ACCCTTTCACTACCCTAACAATCAACCCCTTAACCGATTTTATTCAA
ACCTAACTATGAGACTAACACAACCCAATCATCTTACTCCATCCCATATAATCATCAATTAAACTACCAGTTT
AAACTTGTTCAACCCCCCCCCCACATTCAGAGAAAAAGGAATCAAAC

CTTTATCATCAACTCCCAAAGCTGACATTTTAATTAAACTATTCTCTGATTTATCTCCCCTAAACAGCCCGAA
TAGTCCCCCGAGACAGCCCTCGCACAAGCTCCAAAACCACAAATAAA
GTTAATAACAACCCCCAACCCGCAACAAAAAACTGTCCCACTCCCCAAAAATAAAACATAGCTGCACCATCA
AAATCTAACCGTACAGAAGCCATCCCCTCACAATCCACTGTACCGACA
CCAACCAACCACCCCCCCGGCAACCCAACTAAAAACACCCCACATACTAATACTACAACTAACCCTACAGCA
TATACTAAAACTCGTCAATCCCCCCAAGCCTCAGGAAAAGGATCAGCC
GCCAATGATACTGAGTAAACAAACACTACCAACATTCCGCCTAAATATACTAAAAATAGTACCAACGATACA
AAAGACGATCCTAAACTCATCAACCACCCACACCCCACAATAGACCCC
AACACCAGACCTATCACCCCATAATAAGGAGAAGGATTAGAAGCTACCGCCAATCCTCCTAAAATAAAACA
AATACTTAAAAAAATCATAAAGTATGTCATAGAATTCCTACTTGGTTTC
TCTCCAAGACCTATGGCCTGAAAAACCACCGTTGTAAATTCAACTATAAGAACTTACATTACTACAAACTCC
TAACCAAACACAACCAAAACTTCTTCCACATCACCTTACTTACTACTT
TCAACCTCCCCAACCCATTCTAACAAATGCTTCTTTTTTTTAGACCCATCCAATACAACCAATACCCCTACGT
CAAAAGTACTACAACTATAAAACCAACACACAACTAAA

Part II. Nuclear genome assembly and annotation

Estimates of genome completeness using BUSCO and compleasm, passeriformes lineage.

Completeness of genome assembly based upon BUSCO v5.4.7 analyses:

C:96.2%[S:95.9%,D:0.3%],F:0.6%,M:3.2%,n:10844
10437 Complete BUSCOs (C)
10400 Complete and single-copy BUSCOs (S)
37 Complete and duplicated BUSCOs (D)
62 Fragmented BUSCOs (F)
345 Missing BUSCOs (M)
10844 Total BUSCO groups searched

Completeness of genome assembly based upon compleasm v0.2.2 results:

C:99.45%[S:99.35%, D:0.09%],F:0.30%,M:0.25%,n:10844
10784 Complete BUSCOs (C)
10774 Complete and single-copy BUSCOs (S)
10 Complete and duplicated BUSCOs (D)
33 Fragmented BUSCOs (F)
27 Missing BUSCOs (M)
10844 Total BUSCO groups searched

Completeness of genome assembly based upon combined BUSCO and compleasm analyses:

C:99.72%[S:99.63%,D:0.09%],F:0.-6%,M:0.22%,n:10844
10814 Complete BUSCOs (C)
10804 Complete and single-copy BUSCOs (S)
10 Complete and duplicated BUSCOs (D)
6 Fragmented BUSCOs (F)
24 Missing BUSCOs (M)
10844 Total BUSCO groups searched

BUSCO reported 345 missing, while compleasm found all but 27, of the 10844 passeriformes lineage single copy orthologs. Of the 27 not found by compleasm, 3 of those however were found by BUSCO.

Table S1. *P.nanus* 1.0 Repeat Elements

file name: bPyrNan1.0.fasta			
sequences:	152		
total length:	1072479546 bp (1072468046 bp excl N/X-runs)		
GC level:	42.30%		
bases masked:	120094435 bp (11.20 %)		
	number of elements*	length occupied of sequence	percentage
Retroelements	223983	80090817 bp	7.47%
SINEs:	9529	1115561 bp	0.10%
Penelope:	168	60923 bp	0.01%
LINEs:	180191	58469783 bp	5.45%
CRE/SLACS	0	0 bp	0.00%
L2/CR1/Rex	172099	56473377 bp	5.27%
R1/LOA/Jockey	0	0 bp	0.00%
R2/R4/NeSL	4414	330241 bp	0.03%
RTE/Bov-B	1812	1249014 bp	0.12%
L1/CIN4	78	18471 bp	0.00%
LTR elements:	34263	20505473 bp	1.91%
BEL/Pao	0	0 bp	0.00%
Ty1/Copia	315	128833 bp	0.01%
Gypsy/DIRS1	2077	1374185 bp	0.13%
Retroviral	31601	18935339 bp	1.77%
DNA transposons	26638	5513805 bp	0.51%
hobo-Activator	4104	921735 bp	0.09%
Tc1-IS630-Pogo	570	98656 bp	0.01%
En-Spm	0	0 bp	0.00%
MULE-MuDR	359	165393 bp	0.02%
PiggyBac	0	0 bp	0.00%
Tourist/Harbinger	8018	827140 bp	0.08%
Other (Mirage, P-element, Transib)	0	0 bp	0.00%
Rolling-circles	217	74483 bp	0.01%
Unclassified:	43282	16674470 bp	1.55%
Total interspersed repeats:		102340015 bp	9.54%
Small RNA:	2576	410422 bp	0.04%
Satellites:	3630	1299929 bp	0.12%
Simple repeats:	278611	13149163 bp	1.23%
Low complexity:	54796	3038055 bp	0.28%

Table S2. Scaffold composition: assignment to chromosome and information about size and gene and telomere content.

Scaffold /Chr number	Size bp	Location along genome	Cumulative Genome percentage	BUSCO markers data into categories of total BUSCOs (B), Complete and single-copy (C), Complete and duplicated (D), Fragmented (F), and any duplicated, fragmented or complete (d).	Telomere regions present in the scaffold
Chr1_Lvf	117695848	117,695,848	10.97%	B:890 C:886 F:1 D:2 d:3	telomeres: BOTTOM
Chr2_Lvf	115308024	233,003,872	21.73%	B:904 C:903 F:1 D:0 d:0	telomeres: TOP BOTTOM_near BOTTOM_near
Chr3_Lvf	97861697	330,865,569	30.85%	B:577 C:575 F:0 D:1 d:2	
Chr4_Lvf	74117589	404,983,158	37.76%	B:659 C:658 F:1 D:0 d:0	telomeres: BOTTOM
ChrZ_Lvf	74110192	479,093,350	44.67%	B:536 C:534 F:0 D:1 d:2	telomeres: TOP BOTTOM
Chr6_Lvf	74038366	553,131,716	51.58%	B:548 C:548 F:0 D:0 d:0	telomeres: TOP
Chr7_Lvf	63982733	617,114,449	57.54%	B:728 C:727 F:0 D:1 d:1	telomeres: TOP
Chr8_Lvf	55732288	672,846,737	62.74%	B:404 C:401 F:0 D:2 d:3	telomeres: TOP
Chr9_Lvf	39130240	711,976,977	66.39%	B:374 C:373 F:1 D:0 d:0	
Chr10_Lvf	36877371	748,854,348	69.82%	B:412 C:409 F:1 D:1 d:2	telomeres: BOTTOM
Chr11_Lvf	30701327	779,555,675	72.69%	B:420 C:420 F:0 D:0 d:0	
Chr12_Lvf	25824471	805,380,146	75.10%	B:355 C:355 F:0 D:0 d:0	telomeres: TOP BOTTOM_near BOTTOM
Chr13_Lvf	21584164	826,964,310	77.11%	B:258 C:258 F:0 D:0 d:0	telomeres: TOP
Chr14_Lvf	21327804	848,292,114	79.10%	B:303 C:303 F:0 D:0 d:0	telomeres: TOP
Chr15_Lvf	20799835	869,091,949	81.04%	B:290 C:290 F:0 D:0 d:0	telomeres: TOP
Chr16_Lvf	20111806	889,203,755	82.91%	B:243 C:242 F:0 D:1 d:1	telomeres: TOP TOP_near
Chr17_Lvf	18644349	907,848,104	84.65%	B:251 C:251 F:0 D:0 d:0	telomeres: TOP BOTTOM
Chr18_Lvf	16541591	924,389,695	86.19%	B:318 C:316 F:0 D:1 d:2	telomeres: TOP BOTTOM
Chr19_Lvf	15567225	939,956,920	87.64%	B:262 C:262 F:0 D:0 d:0	telomeres: TOP BOTTOM
Chr20_Lvf	15002696	954,959,616	89.04%	B:274 C:274 F:0 D:0 d:0	telomeres: TOP
Chr21_Lvf	12777097	967,736,713	90.23%	B:229 C:229 F:0 D:0 d:0	telomeres: TOP TOP_near MIDDLE MIDDLE
Chr22_Lvf	12573319	980,310,032	91.41%	B:223 C:222 F:1 D:0 d:0	telomeres: MIDDLE
Chr23_Lvf	11756239	992,066,271	92.50%	B:248 C:248 F:0 D:0 d:0	telomeres: TOP
Chr24_Lvf	8003886	1,000,070,157	93.25%	B:188 C:188 F:0 D:0 d:0	telomeres: TOP BOTTOM
Chr25_Lvf	7744152	1,007,814,309	93.97%	B:139 C:139 F:0 D:0 d:0	telomeres: TOP BOTTOM
Chr26_Lvf	7268963	1,015,083,272	94.65%	B:134 C:134 F:0 D:0 d:0	telomeres: BOTTOM
Chr27_Lvf	6803409	1,021,886,681	95.28%	B:154 C:153 F:0 D:1 d:1	telomeres: TOP
Chr28_Lvf	6444326	1,028,331,007	95.88%	B:170 C:170 F:0 D:0 d:0	telomeres: TOP
Chr29_Lvf	6179786	1,034,510,793	96.46%	B:164 C:163 F:0 D:1 d:1	telomeres: TOP
Chr30_Lvf	4788275	1,039,299,068	96.91%	B:85 C:82 F:0 D:2 d:3	telomeres: TOP BOTTOM
Chr31_Lvf	4683349	1,043,982,417	97.34%		telomeres: MIDDLE BOTTOM
Chr32_Lvf	4644022	1,048,626,439	97.78%	B:61 C:60 F:0 D:1 d:1	telomeres: TOP BOTTOM
Chr33_Lvf	2372421	1,050,998,860	98.00%	B:29 C:29 F:0 D:0 d:0	telomeres: BOTTOM
Chr34_Lvf	2292617	1,053,291,477	98.21%		telomeres: TOP TOP_near
Chr35_Lvf	1968829	1,055,260,306	98.39%	B:3 C:0 F:0 D:1 d:3 *	telomeres: MIDDLE BOTTOM
Chr36_Lvf	1929877	1,057,190,183	98.57%		telomeres: TOP
Chr37_Lvf	1896899	1,059,087,082	98.75%		telomeres: TOP
Chr38_Lvf	1688031	1,060,775,113	98.91%		
scaffold_39	792425	1,061,567,538	98.98%		
scaffold_40	681081	1,062,248,619	99.05%		

scaffold_47	617421	1,062,866,040	99.10%		
scaffold_41	466723	1,063,332,763	99.15%		
scaffold_42	392199	1,063,724,962	99.18%		
scaffold_43	352952	1,064,077,914	99.22%		telomeres: TOP
scaffold_44	352859	1,064,430,773	99.25%		
scaffold_45	332306	1,064,763,079	99.28%		telomeres: TOP
scaffold_46	331746	1,065,094,825	99.31%	B:1 C:1 F:0 D:0 d:0	telomeres: TOP
scaffold_48	274237	1,065,369,062	99.34%		
scaffold_49	255847	1,065,624,909	99.36%		
scaffold_50	232336	1,065,857,245	99.38%		
scaffold_51	222529	1,066,079,774	99.40%		
scaffold_52	212371	1,066,292,145	99.42%		
scaffold_53	208070	1,066,500,215	99.44%		
scaffold_54	195784	1,066,695,999	99.46%		telomeres: TOP
scaffold_55	183290	1,066,879,289	99.48%		telomeres: TOP
scaffold_56	178647	1,067,057,936	99.49%		
scaffold_57	174567	1,067,232,503	99.51%		telomeres: BOTTOM
scaffold_58	171796	1,067,404,299	99.53%		
scaffold_59	147170	1,067,551,469	99.54%		telomeres: TOP
scaffold_60	140188	1,067,691,657	99.55%		
scaffold_61	137719	1,067,829,376	99.57%		
scaffold_62	134803	1,067,964,179	99.58%		
scaffold_63	131517	1,068,095,696	99.59%		
scaffold_64	127349	1,068,223,045	99.60%	B:1 C:1 F:0 D:0 d:0	
scaffold_65	126107	1,068,349,152	99.61%		
scaffold_66	110870	1,068,460,022	99.63%		
scaffold_67	109590	1,068,569,612	99.64%		
scaffold_68	105808	1,068,675,420	99.65%		
scaffold_69	101436	1,068,776,856	99.65%		
scaffold_70	98449	1,068,875,305	99.66%		
scaffold_71	91546	1,068,966,851	99.67%		
scaffold_72	81735	1,069,048,586	99.68%		
scaffold_74	81294	1,069,129,880	99.69%		
scaffold_73	81234	1,069,211,114	99.70%		
scaffold_75	81173	1,069,292,287	99.70%		
scaffold_76	81135	1,069,373,422	99.71%		
scaffold_77	78901	1,069,452,323	99.72%		
scaffold_78	78067	1,069,530,390	99.73%		
scaffold_79	72673	1,069,603,063	99.73%		
scaffold_80	72651	1,069,675,714	99.74%		
scaffold_81	72397	1,069,748,111	99.75%		
scaffold_82	69919	1,069,818,030	99.75%		
scaffold_83	68207	1,069,886,237	99.76%		
scaffold_84	68123	1,069,954,360	99.76%		
scaffold_85	66988	1,070,021,348	99.77%		
scaffold_86	66445	1,070,087,793	99.78%		
scaffold_87	65391	1,070,153,184	99.78%		
scaffold_88	65280	1,070,218,464	99.79%		
scaffold_89	64227	1,070,282,691	99.80%		

scaffold_90	62549	1,070,345,240	99.80%		
scaffold_91	61490	1,070,406,730	99.81%		
scaffold_92	60214	1,070,466,944	99.81%		
scaffold_93	57685	1,070,524,629	99.82%		
scaffold_94	55126	1,070,579,755	99.82%		
scaffold_95	54002	1,070,633,757	99.83%		
scaffold_96	53407	1,070,687,164	99.83%		
scaffold_97	53309	1,070,740,473	99.84%		
scaffold_98	53113	1,070,793,586	99.84%		
scaffold_99	51578	1,070,845,164	99.85%		
scaffold_100	51029	1,070,896,193	99.85%		
scaffold_101	49169	1,070,945,362	99.86%		
scaffold_102	47514	1,070,992,876	99.86%		
scaffold_103	47339	1,071,040,215	99.87%		
scaffold_104	46409	1,071,086,624	99.87%		
scaffold_105	45961	1,071,132,585	99.87%		
scaffold_106	44915	1,071,177,500	99.88%		
scaffold_107	43785	1,071,221,285	99.88%		
scaffold_108	42595	1,071,263,880	99.89%		
scaffold_109	42074	1,071,305,954	99.89%		
scaffold_110	39410	1,071,345,364	99.89%		
scaffold_111	38733	1,071,384,097	99.90%		
scaffold_112	37418	1,071,421,515	99.90%		
scaffold_113	36837	1,071,458,352	99.90%		
scaffold_114	36251	1,071,494,603	99.91%		
scaffold_115	35676	1,071,530,279	99.91%		
scaffold_116	34611	1,071,564,890	99.91%		
scaffold_117	34461	1,071,599,351	99.92%		
scaffold_118	33737	1,071,633,088	99.92%		
scaffold_119	33636	1,071,666,724	99.92%		
scaffold_120	33209	1,071,699,933	99.93%		
scaffold_121	32981	1,071,732,914	99.93%		
scaffold_122	31952	1,071,764,866	99.93%		
scaffold_123	31535	1,071,796,401	99.94%		
scaffold_124	30636	1,071,827,037	99.94%		telomeres: TOP
scaffold_125	30585	1,071,857,622	99.94%		
scaffold_126	30321	1,071,887,943	99.94%		
scaffold_127	30188	1,071,918,131	99.95%		
scaffold_128	29982	1,071,948,113	99.95%		
scaffold_129	29297	1,071,977,410	99.95%		telomeres: TOP_near
scaffold_130	29174	1,072,006,584	99.96%		
scaffold_131	28482	1,072,035,066	99.96%		
scaffold_132	27437	1,072,062,503	99.96%		
scaffold_133	26543	1,072,089,046	99.96%		
scaffold_134	25991	1,072,115,037	99.97%		
scaffold_135	25876	1,072,140,913	99.97%		
scaffold_136	24322	1,072,165,235	99.97%		
scaffold_137	23071	1,072,188,306	99.97%		telomeres: TOP_near
scaffold_138	22816	1,072,211,122	99.97%		telomeres: TOP

scaffold_139	22711	1,072,233,833	99.98%		
scaffold_140	21735	1,072,255,568	99.98%		
scaffold_141	20752	1,072,276,320	99.98%		telomeres: TOP_near
scaffold_142	20686	1,072,297,006	99.98%		
scaffold_143	20552	1,072,317,558	99.98%		
scaffold_144	20471	1,072,338,029	99.99%		
scaffold_145	20400	1,072,358,429	99.99%		
scaffold_146	19697	1,072,378,126	99.99%		
scaffold_147	18830	1,072,396,956	99.99%		
scaffold_148	18614	1,072,415,570	99.99%		
scaffold_149	18076	1,072,433,646	100%		
scaffold_150	17734	1,072,451,380	100%		telomeres: TOP_near
scaffold_151	15505	1,072,466,885	100%		telomeres: TOP_near
scaffold_152	12661	1,072,479,546	100%		

Candidate Annotation summary stats:

30101 Genes
13.26% Percentage of Genome
142232926 Total Gene Length
1072479546 Assembly Length

31748 mRNA
5119.83 Mean Length
171634 Longest
201 Shortest
162544457 Total mRNA Length

161303 Exons
5.08 Mean per mRNA
199.33 Mean Length
1012.74 Mean Length per mRNA
13482 Longest Exon
3 Shortest Exon
32152359 Total Exon Length
5056 15.93% Single Exon mRNA

129555 Introns
4.08 Mean per mRNA
993.75 Mean Length
4055.21 Mean Length per mRNA
26550 Longest Intron
41 Shortest Intron
128744773 Total Intron Length

15.16% mRNA
3.00% Exons
12.00% Introns

1072479546 Assembly Length

16468 Genes with gene names
17902 mRNA with gene names
19303 mRNA with gene descriptions

Candidate Annotation associated files:

bPyrNan1.0.basic_gff_stats.txt: Contains basic statistics regarding the annotation.
bPyrNan1.0.codingseq.fna: is a fasta file containing nucleotide sequence for each annotated coding sequence.
bPyrNan1.0.faa: is a fasta file containing amino acid sequence for each annotated coding sequence.
bPyrNan1.0.gff: is a general feature format of annotations for the genome.

Table S3. Statistics of nine assembled genomes from birds of the family Tyranidae, including the genome of *Pyrocephalus nanus*.

Species	Date	Genome size	# Scaffold	Scaffold N50	# Contigs	Contig N50	Assembly level	Sequence platform
Empidonax traillii	3-Apr-2018	1.1 Gb	7791	895.1 kb	45,995	86.6 kb	Scaffold	Illumina HiSeq
Myiozetetes cayanensis	10-Mar-2022	1.1 GB	1692	63.7 Mb	1,873	13 Mb	Scaffold	Oxford nanopore; Illumina
Tyrannus savana	10-Jul-2020	1.1 GB	35393	436.2 kb	70,134	72.3 kb	Scaffold	Illumina HiSeq
Neopipo cinnamomea	10-Jul-2020	1 GB	11198	486 kb	61,449	43.1 kb	Scaffold	Illumina HiSeq
Tyrannus tyrannus	9-Dec-2022	1.1 GB	43947	63.1 Mb	114,348	22.5 kb	Scaffold	Illumina
Pitangus sulphuratus	20-Mar-2023	1 GB	11553	829.5 kb	116,115	15.7 kb	Scaffold	Illumina MiSeq; Illumina HiSeq
Empidonax alnorum	2-Jun-2022	1.1 GB	15260	7.7 Mb	31,625	165.8 kb	Scaffold	Illumina HiSeq
Mionectes macconnelli	10-Jul-2020	1 GB	4545	1.6 Mb	33,645	72.8 kb	Scaffold	Illumina HiSeq
Pyrocephalus nanus	16-Nov-2023	1.07 GB	152	74.0 Mb	267	17.8 Mb	Chromosome & scaffold	PacBio Hifi, Illumina HiC

References

Allio R, et al. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol Ecol Resour. 20(4):892-905.

Bernt M, et al. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 69(2):313-319.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772-780.

CHAPTER 3

Phylogeny, population structure, and conservation genomics of Galapagos vermilion flycatchers (genus *Pyrocephalus*)

DAVID J. ANCHUNDIA · ATHENA W. LAM · JAMES B. HENDERSON · MATTHEW H. VAN DAM · COURTNEY L. PIKE · SABINE TEBBICH · BIRGIT FESSL · HEINZ RICHNER · JOHN P. DUMBACHER

Photo by David Anchundia

Submitted 12 November 2024 to journal Molecular Phylogenic and Evolution, in revision. Preprint http://dx.doi.org/10.2139/ssrn.5026862

Molecular Phylogenetics and Evolution

Phylogeny, population structure, and conservation genomics of Galapagos vermilion flycatchers (genus Pyrocephalus)

--Manuscript Draft--

Manuscript Number:	MPE-D-24-00506
Article Type:	Research Paper
Keywords:	Island biogeography; Pyrocephalus nanus; P. dubius; Pyrocephalus radiation; bioacoustics; whole genome sequence
Corresponding Author:	David Anchundia University of Vienna Department of Behavioural and Cognitive Biology AUSTRIA
First Author:	David Anchundia
Order of Authors:	David Anchundia
	Athena W. Lam, Phd
	James B. Henderson
	Matthew H. Van Dam, Phd
	Courtney L. Pike, Phd
	Sabine Tebbich, Phd
	Birgit Fessl, Phd
	Heinz Richner, Phd
	John P. Dumbacher, Phd
Abstract:	The Galapagos archipelago is recognized for its endemism produced by its isolation in the Tropical Pacific Ocean as well as the differentiation of endemic groups across islands. However, there is still a lack of knowledge about many groups, such as the genus Pyrocephalus, a group of flycatchers whose populations have declined rapidly in recent decades. Using samples from 53 Pyrocephalus individuals from 11 Galapagos islands, we reconstructed a phylogeny using 233,916 SNPs obtained by sequencing whole-genomes. Pyrocephalus in Galapagos is much older than previously thought, with an estimated arrival of 1.3 million years ago (MYA). Our data indicates that there are at least seven genetically distinct groups in the Galapagos, including Pyrocephalus dubius, a recognized species endemic to San Cristobal Island, at the base of the topology; however, this species is likely already extinct. Pyrocephalus populations on Floreana Island also diverged long ago and had been isolated for at least 1 MYA, but are also presumed extinct. Also, we identified different divergent groups of Pyrocephalus populations from the islands Isabela-Fernandina, Pinta, Santa Cruz, Pinzón-Rabida and Santiago-Marchena that have remained isolated between 0.77 and 0.15 MYA. We compared the morphology of P. nanus populations among island populations using wing size and found significant differences. Birds from Pinta and Isabela had the largest wing size, while those with smaller wing sizes were birds from Pinzon and Santa Cruz. Bioacoustic analyses among these genetic lineages confirmed that there are significant differences in male territorial song. These results suggest that these populations have had little interaction or migration recently. Furthermore, we evaluated whether the Pyrocephalus population on Santa Cruz Island is suffering from inbreeding due to its small population size by comparing modern samples with museum samples collected from Santa Cruz in the past (60-120 yrs ago), when the populations were larger. We found a significant reduction in nucleotide diversity within the island. However, we found no evidence of breeding among close relatives, suggesting that the population may have declined rapidly.

Phylogeny, population structure, and conservation genomics of Galapagos vermilion flycatchers (genus *Pyrocephalus*)

David J. Anchundia[1,2,3*], Athena W. Lam[3], James B. Henderson[3], Matthew H. Van Dam[3], Courtney L. Pike[1,2], Sabine Tebbich[1], Birgit Fessl[2], Heinz Richner[4], John P. Dumbacher[3]

[1]University of Vienna, Department of Behavioural and Cognitive Biology, Vienna, Austria

[2]Charles Darwin Research Station, Charles Darwin Foundation, Santa Cruz, Galápagos, Ecuador

[3]California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA 94118, USA

[4]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

*Author for correspondence; email: davidanchundia@gmail.com

Abstract

The Galapagos archipelago is recognized for its endemism produced by its isolation in the Tropical Pacific Ocean as well as the differentiation of endemic groups across islands. However, there is still a lack of knowledge about many groups, such as the genus *Pyrocephalus*, a group of flycatchers whose populations have declined rapidly in recent decades. Using samples from 53 *Pyrocephalus* individuals from 11 Galapagos islands, we reconstructed a phylogeny using 233,916 SNPs obtained by sequencing whole-genomes. *Pyrocephalus* in Galapagos is much older than previously thought, with an estimated arrival of 1.3 million years ago (MYA). Our data indicates that there are at least seven genetically distinct groups in the Galapagos, including *Pyrocephalus dubius*, a recognized species endemic to San Cristobal Island, at the base of the topology; however, this species is likely already extinct. *Pyrocephalus* populations on Floreana Island also diverged long ago and had been isolated for at least 1 MYA, but are also presumed extinct. Also, we identified different divergent groups of *Pyrocephalus* populations from the islands Isabela-Fernandina, Pinta, Santa Cruz, Pinzón-Rabida and Santiago-Marchena that have remained isolated between 0.77 and 0.15 MYA.

We compared the morphology of *P. nanus* populations among island populations using wing size and found significant differences. Birds from Pinta and Isabela had the largest wing size, while those with smaller wing sizes were birds from Pinzon and Santa Cruz. Bioacoustic analyses among these genetic lineages confirmed that there are significant differences in male territorial song. These results suggest that these populations have had little interaction or migration recently. Furthermore, we evaluated whether the *Pyrocephalus* population on Santa Cruz Island is suffering from inbreeding due to its small population size by comparing modern samples with museum samples collected from Santa Cruz in the past (60-120 yrs ago), when the populations were larger. We found a significant reduction in nucleotide diversity within the island. However, we found no evidence of breeding among close relatives, suggesting that the population may have declined rapidly.

Keywords
Island biogeography, *Pyrocephalus nanus*, *P. dubius*, *Pyrocephalus* radiation, bioacoustics, whole genome sequence

1. Introduction

Biodiversity in the Galapagos Islands is in decline (Benitez-Capistros et al., 2014), thus it is important to investigate poorly-studied species in order to understand their status or need for protection or conservation. Currently, several species of landbirds in the Galapagos Islands are declining (Dvorak et al., 2012, 2017, 2019; Fessl et al., 2017), including the endemic Little Vermilion Flycatcher, *Pyrocephalus nanus* (BirdLife International 2023). This species was formerly distributed across at least twelve islands in the Galapagos (Gifford, 1919; Swarth, 1931) and is currently among the avian species showing the greatest population decline (Merlen 2013; Fessl et al., 2017; Leuba et al., 2020; Mosquera et al., 2022; Anchundia et al., 2024a). The closely related Least Vermilion Flycatcher (*P. dubius*) was categorized as an endemic species after its presumed extinction on San Cristobal Island (Carmi et al., 2016; Del Hoyo et al., 2016; Chesser et al., 2022) and there have been no confirmed sightings since 2008 (Dvorak et al., 2017). These species are also the most brightly-colored and arguably the most charismatic landbird species in the Galapagos. Despite severe population reductions, relatively little attention has been dedicated to the population genetics of this genus in the Galapagos. To effectively conserve the species, including translocating individuals and augmenting declining populations, practitioners need to understand genetic relationships among islands' populations.

Since the discovery of the vermilion flycatcher in Galapagos by Charles Darwin in 1835 (Gould, 1841), their populations across the Galapagos archipelago have been the subject of taxonomic controversy. Several authors have separated them into two to five species and two subspecies based on morphological characters, with the authors disagreeing on how many species exist or where the divisions occur (Gould, 1841; Ridgway, 1894; Rothschild & Hartert, 1899; Snodgrass & Heller, 1904; Swarth, 1931; Carmi et al., 2016). Thus, *Pyrocephalus* appears to be a cryptic species complex in which several species may exist, but their morphological differences are subtle and their taxonomic or ecological implications unclear, and thus they have been difficult to classify (Bickford, 2007, Mills et al., 2017).

However, analysis using mitochondrial genes confirmed the existence of at least two species, corresponding to *P. dubius* (San Cristobal Island) and *P. nanus* (the rest of the Galapagos; Carmi et al., 2016). Additionally, this study found distinct genetic groups within *P. nanus* (Carmi et al., 2016) suggesting that there are potentially genetically distinct populations, subspecies, or cryptic species that have not been described. The conclusions of Carmi et al. (2016), were based primarily on only a few

mitochondrial DNA sequences under 1000 bp derived from museum specimens. Therefore, further study is needed and should include greater genetic and geographic sampling to understand population structure and taxonomy of *Pyrocephalus* in the Galapagos (Ennos et al., 2005; Irestedt et al., 2017). To help corroborate genetic data, we would like to also evaluate morphological and bioacoustic data among *P. nanus* populations (Carmi et al., 2016). Song variation among populations often reflects reproductive isolation (Seddon & Tobias, 2007) and both birders and guides have reported song differences among island populations.

The decline of Galapagos *Pyrocephalus* is linked to human impacts across many of the Galapagos Islands (Campbell & Donlan, 2005, Watson et al., 2010, Dvorak et al., 2019). This has led to the apparent extinction of *P. dubius* and the extirpation of *P. nanus* from Floreana, Santa Fe, and Baltra islands. Additionally, *P. nanus* is becoming rare on Santa Cruz, Santiago, and Rabida islands (Fessl et al., 2017; Harris, 1973; Anchundia pers. obs). Currently, the only islands with stable populations are Isabela, Pinzon, Pinta, Marchena, and Fernandina (Charles Darwin Foundation, unpublished data). On Santa Cruz Island, where *P. nanus* was abundant 30 years ago, the population has decreased to fewer than 30 to 40 individuals occurring only in the highlands, with an apparent sex-ratio of approximately two males to each female (Anchundia et al., 2024a). These declines and biased sex-ratio can lead to a reduced effective population size and a genetic bottleneck. This dramatic decline on Santa Cruz has been linked to habitat degradation due to invasive blackberry (*Rubus niveus*) as well as reproductive losses from the invasive parasitic avian vampire fly (*Philornis downsi*), whose larvae parasitize nestlings and decrease host breeding success (Leuba et al., 2020; Mosquera et al., 2022; Anchundia et al., 2024a).

The potential genetic bottleneck of *P. nanus* in Santa Cruz is a major concern. Small populations in isolated localities are vulnerable to stochastic environmental and demographic events that accelerate extinction (Lande, 1993; Keller & Waller, 2002). Sudden variation in climate, food, predators, competition, and parasites make population trajectories unpredictable (Lande, 1993; Keller & Waller, 2002). When populations become small, like those in Santa Cruz, they can lose alleles by random drift and lead to inbreeding depression, fitness declines, and ultimately the inability to reproduce and survive (Reed & Frankham, 2003; Brekke et al., 2010, Duntsch et al., 2023). Signs of inbreeding depression in avian populations are well documented, including a decrease in egg fertility (Jamieson & Ryan, 2000), reduction in egg hatching (Briskie & Machiston, 2004; Kruuk et al., 2002; Jamieson et al., 2003; Heber & Briskie, 2010; Greenwood et al., 1978), and lower weight and smaller size upon hatching (Keller, 1998). Nestlings with greater inbreeding have lower probabilities of survival (Greenwood et al., 1978; Van Noordwijk & Scharloo, 1981; Keller, 1998; Kruuk et al., 2002).

Inbreeding also decreases the genetic diversity that allows populations to adapt and evolve (Loeschcke et al., 2013; Willi et al., 2022; Duntsch et al., 2023). Because *P. nanus* on some other

islands show similar population size reductions, loss of genetic diversity is a concern for Galapagos *Pyrocephalus* conservation. Most populations lack the baseline demographic data required to assess their population status. With a strong genetic dataset, we can estimate genetic demography parameters, estimate potential inbreeding, search for evidence of consanguineous mating, and evaluate whether they are losing genetic diversity (Willi et al., 2022). This information is valuable for informing future recovery programs of *Pyrocephalus* on the Galapagos.

Our study aims to better understand the phylogenetic relationships, population genetics, and connections among the different populations of *P. nanus* in the Galapagos using whole sequencing genome. The first aim is to reconstruct a well-supported phylogeny of all island populations of *Pyrocephalus* and assess whether there are distinct genetic and taxonomic groups within *P. nanus.* We first briefly assess available bioacoustics and morphological data among islands for concordance with genetic data and recommend revisions in its taxonomy where appropriate. Second, we aim to assess the current nucleotide diversity, heterozygosity and consanguineous mating of *P. nanus* on Santa Cruz Island by comparing present-day genetic variation across thousands of single nucleotide polymorphisms (SNPs) to the variation present in museum specimens collected from Santa Cruz between 60 to 115 years ago. Museum specimens utilized for this study were collected at different sites on Santa Cruz island when *P. nanus* was apparently abundant, which will serve as a baseline. Third, we aim to evaluate the current nucleotide diversity and heterozygosity of the populations of Fernandina, Isabela, Pinta, and Pinzon islands, using modern samples.

2. Methods

2.1. Study site and study species

This study was conducted on the Galapagos archipelago, an isolated volcanic island system located in the equatorial eastern tropical Pacific about 1000 km west of mainland Ecuador (Fig. 1). *Pyrocephalus* species have marked sexual dimorphism. Adult males have red plumage on the chest and crown of the head and glossy black plumage on the back, tail, and wings, while females have more cryptic pale yellow chest and crown and brown back, wing, and tail feathers.

Pyrocephalus nanus is significantly larger than the closely related species *P. dubius*, with an average wing length of 63.4 mm (range 61 - 66 mm) and average tail length of 52.7 mm (range 50 - 56 mm) (Snodgrass & Heller, 1904), while *P. dubius* has average wing length of 57.1 mm (range 55 - 59 mm) and average tail length of 48.9 mm (range 48 - 52 mm) (Snodgrass & Heller, 1904). The smaller size of *P. dubius* was among the reasons cited by several authors who separated it into a different taxon even before the rise of genetic analysis (Gould, 1841; Ridgway, 1894; Rothschild & Hartert, 1899; Snodgrass & Heller, 1904; Swarth, 1931).

2.2. Genetic Sampling

2.2.1. Modern samples

Approximately 50 µl of blood was sampled from the brachial vein of the birds captured in mist nets. Birds were sampled from the Santa Cruz, Pinzon, Pinta, and Fernandina islands and from three of the six volcanoes of Isabela Island (Sierra Negra, Alcedo and Wolf volcanoes). Due to the remoteness of these islands and without immediate access to refrigeration, the samples were preserved in Longmire buffer (Longmire et al., 1997) to prevent degradation. Subsequently, upon arrival at the Charles Darwin Research Station on Santa Cruz Island, the samples were stored in a freezer at -27°C to prevent degradation. An attempt was made to acquire samples from Santiago and Rabida islands where this species is apparently still present, but no birds were seen or captured during six days of fieldwork on these islands. Samples were collected on Marchena Island, but due to movement restrictions during the Covid pandemic, these samples could not be exported for laboratory analysis. Frozen tissue samples from *P. obscurus* from Arizona, USA (Burke Museum Collection, Table 1) were used as an outgroup taxon for the Galapagos *Pyrocephalus* group.

2.2.2. Ancient samples

On islands where modern samples of *P. nanus* could not be obtained (Floreana, Rabida, Santiago, Marchena, Baltra, and Wolf), DNA was extracted from museum specimens of the California Academy of Sciences (CAS) (Table 1). Also, samples of *P. dubius* from San Cristobal (Table 1) were added to examine the relationships among these species. Small skin samples were shaved off from the toe pad of these specimens using a scalpel, yielding approximately 1 mm^3 of tissue (Fig. 1).

Table 1

List of samples used in this study. Sixteen bird skins considered historical and preserved in the CAS museum were used. The skins were collected between 1899 and 1962. One sample from Arizona and 38 modern blood samples collected between 2014 and 2021 were used. In the column "Analysis performed with sample", PHY=phylogenetic, STR=structure and PGEN=population genetics. All samples are uploaded on NCBI, SRA ascension project PRJNA1064398.

Island	*Type of sample Ancient (A) Modern (M)*	*Locality*	*Sex*	*Date*	*CAS Museum No. / Sample ID*	*Analysis performed with sample*
Baltra	*Toe pad skin museum (A)*	*Arid zone*	*F*	*21 Nov 1905*	*ORN3211 / 1BLT-39A*	*PGEN*
Fernandina	*Toe pad skin museum (A)*	*---*	*F*	*5 Apr 1899*	*ORN76696 / 1FRN-23A*	*PHY/STR/PGEN*
Fernandina	*Blood sample (M)*	*Volcano rim*	*M*	*21 Feb 2021*	*2FRN-28BL*	*PHY/STR/PGEN*
Fernandina	*Blood sample (M)*	*Volcano rim*	*F*	*21 Feb 2021*	*3FRN-29BL*	*PHY/STR/PGEN*
Fernandina	*Blood sample (M)*	*Volcano rim*	*F*	*22 Feb 2021*	*4FRN-30BL*	*PHY/STR/PGEN*
Fernandina	*Blood sample (M)*	*Volcano rim*	*M*	*24 Feb 2021*	*5FRN-31BL*	*PHY/STR/PGEN*
Floreana	*Toe pad skin museum (A)*	*Highland Wittmer farm*	*M*	*24 Jan 1962*	*ORN86222 / 1FLO-43A*	*PHY/STR*
Floreana	*Toe pad skin museum (A)*	*Highland Wittmer farm*	*F*	*24 Jan 1962*	*ORN86223 / 2FLO-27A*	*PHY/STR*
Floreana	*Toe pad skin museum (A)*	*---*	*F*	*10 May 1899*	*ORN76686 / 3FLO-14A*	*PHY/STR*
Isabela	*Blood sample (M)*	*Alcedo Volcano rim*	*M*	*26 Jan 2017*	*1ISB-2BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Alcedo Volcano rim*	*F*	*26 Jan 2017*	*2ISB-3BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*F*	*19 Nov 2014*	*3ISB-210BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*F*	*21 Nov 2014*	*4ISB-211BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*M*	*26 Nov 2014*	*5ISB-213BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*M*	*26 Nov 2014*	*6ISB-214BL*	*PHY/STR/PGEN*

Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*F*	*26 Nov 2014*	*7ISB-215BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*M*	*27 Nov 2014*	*8ISB-218BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*M*	*28 Nov 2014*	*9ISB-219BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*M*	*18 Nov 2014*	*10ISB-207BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*F*	*27 Nov 2014*	*11ISB-216BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Sierra Negra volcano rim*	*F*	*27 Nov 2014*	*12ISB-217BL*	*PHY/STR/PGEN*
Isabela	*Blood sample (M)*	*Wolf volcano slope*	*M*	*28 Jan 2020*	*13ISB-17BL*	*PHY/STR/PGEN*
Marchena	*Toe pad skin museum (A)*	*----*	*F*	*17 Sept 1906*	*ORN3206 / 1MRC-41A*	*PHY/STR*
Marchena	*Toe pad skin museum (A)*	*----*	*F*	*17 Sept 1906*	*ORN3199 / 2MRC-20A*	*PHY/STR*
Pinta	*Toe pad skin museum (A)*	*----*	*M*	*25 Jun 1899*	*ORN76689 / 1PNT-19A*	*PHY/STR*
Pinta	*Blood sample (M)*	*Highland near submit*	*M*	*11 Feb 2017*	*2PNT-4BL*	*PHY/STR/PGEN*
Pinta	*Blood sample (M)*	*Highland near submit*	*F*	*11 Feb 2017*	*3PNT-5BL*	*PHY/STR/PGEN*
Pinta	*Blood sample (M)*	*Transition zone*	*F*	*31 Jan 2019*	*4PNT-6BL*	*PHY/STR/PGEN*
Pinta	*Blood sample (M)*	*Transition zone*	*M*	*31 Jan 2019*	*5PNT-7BL*	*PHY/STR/PGEN*
Pinzon	*Blood sample (M)*	*Arid zone (top of island)*	*F*	*18 May 2019*	*1PNZ-10BL*	*PHY/STR/PGEN*
Pinzon	*Blood sample (M)*	*Arid zone (top of island)*	*F*	*6 March 2015*	*2PNZ-411BL*	*PHY/STR/PGEN*
Pinzon	*Blood sample (M)*	*Arid zone (near the coast)*	*F*	*21 May 2019*	*3PNZ-11BL*	*PHY/STR/PGEN*
Pinzon	*Blood sample (M)*	*Arid zone (near the coast)*	*F*	*21 May 2019*	*4PNZ-12BL*	*PHY/STR/PGEN*
Pinzon	*Blood sample (M)*	*Arid zone (near the coast)*	*F*	*3 March 2015*	*5PNZ-407BL*	*PHY/STR/PGEN*
Pinzon	*Blood sample (M)*	*Arid zone (near the coast)*	*F*	*5 March 2015*	*6PNZ-408BL*	*PHY/STR/PGEN*
Rabida	*Toe pad skin museum (A)*	*----*	*M*	*21 Dec 1905*	*ORN3112 / 1RBD-37A*	*PHY*
San Cristobal	*Toe pad skin museum (A)*	*----*	*M*	*17 Oct 1905*	*ORN3287 / 1SCB-1A*	*PHY*
Santa Cruz	*Toe pad skin museum (A)*	*Transition zone*	*F*	*22 Nov 1961*	*ORN86209 / 1SCZ-2A*	*PHY/STR/PGEN*
Santa Cruz	*Toe pad skin museum (A)*	*Transition zone*	*M*	*25 Nov 1961*	*ORN86210 / 2SCZ-15A*	*PHY/STR/PGEN*
Santa Cruz	*Toe pad skin museum (A)*	*Arid zone*	*M*	*19 Nov 1905*	*ORN3087 / 3SCZ-42A*	*PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*M*	*15 Mar 2019*	*4SCZ-8BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*M*	*19 Dec 2019*	*5SCZ-13BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*F*	*1 Sept 2020*	*6SCZ-21BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*M*	*4 Dec 2020*	*7SCZ-27BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*M*	*22 Mar 2021*	*8SCZ-32BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*F*	*6 Feb 2020*	*9SCZ-18BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*F*	*5 Mar 2020*	*10SCZ-19BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*F*	*1 Sept 2020*	*11SCZ-22BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*F*	*19 Dec 2019*	*12SCZ-14BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*M*	*19 Dec 2019*	*13SCZ-15BL*	*PHY/STR/PGEN*
Santa Cruz	*Blood sample (M)*	*Highland (Mina Roja)*	*M*	*6 Dec 2016*	*14SCZ-1BL*	*PHY/STR/PGEN*
Santiago	*Toe pad skin museum (A)*	*----*	*F*	*19 Dec 1905*	*ORN3233 / 1STG-40A*	*PHY/STR*
Santiago	*Toe pad skin museum (A)*	*----*	*M*	*23 Dec 1905*	*ORN3067 / 2STG-21A*	*PHY/STR*
Wolf	*Toe pad skin museum (A)*	*----*	*F*	*24 Sept 1906*	*ORN3185 / 1WLF-38A*	*PHY*
USA-Arizona	*Frozen tissue (M)*	*Duncan*	*M*	*15 Sept 2003*	*Burke Museum 77983 / 1ARZ-ARZ*	*PHY*

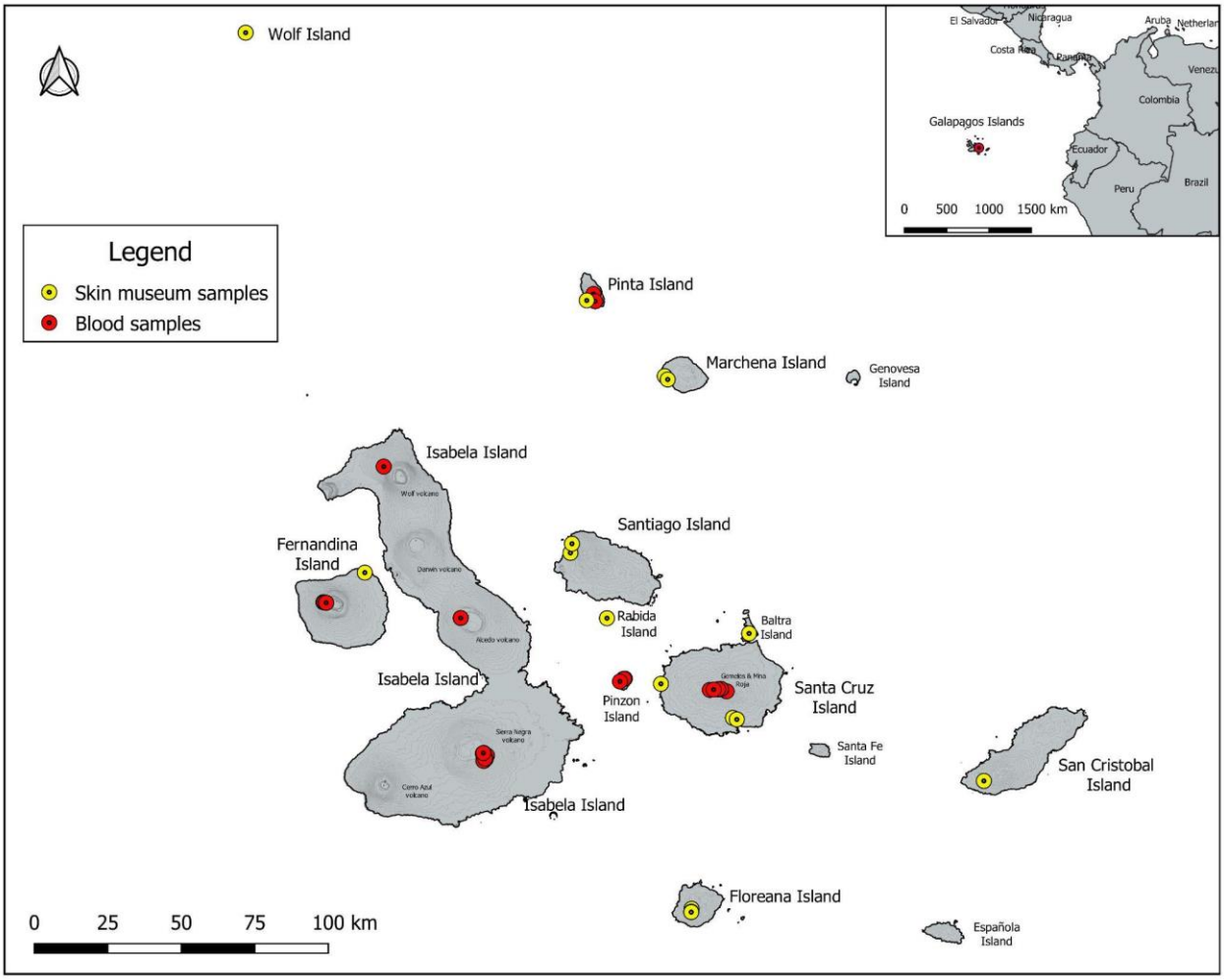


Fig. 1. Map of the Galapagos Islands. Red dots indicate localities where modern samples were collected and yellow dots indicate the approximate locations where the museum skins were collected.

2.3. Phylogenetics

For phylogenetic analysis, we used 53 samples (Table 1; 38 modern and 15 museum samples). For outgroups, one sample was used from San Cristobal *P. dubius* and one sample from *P. obscurus* from Duncan, Arizona (USA) (latitude 32.9, longitude -109.22). Santa Fe Island is the only historically documented island population (last documented in 1929) not represented in our analysis (Fisher & Wetmore, 1931).

2.3.1. DNA extraction of modern samples

DNA was extracted using the Monarch® Genomic DNA Purification Kit following the manufacturer's protocol from New England Biolab (NEB). DNA was quantified using a Qubit 3.0 (Invitrogen). All modern samples had DNA concentrations greater than 100 ng/µl and fragment sizes ~5000 bp long, confirming that samples were well preserved in Longmire buffer without refrigeration for several days in the field. We sheared long fragments into smaller fragments for library construction using a *qSonica* sonicator.

2.3.2. DNA extraction of ancient samples

For museum skin DNA extraction, we used a dedicated ancient DNA laboratory at CAS to minimize the risk of contamination. For the extraction, the phenol-chloroform-isoamyl alcohol method was used, followed by concentration using gravity-assisted dialysis in Centricon (Ultracel YM-30) spin columns. The samples were quantified on a Qubit after extraction and the DNA concentration was typically much lower (< ~5 ng/µl) and more degraded (average fragment sizes of ~120 bp or less) than modern samples.

2.3.3. Sequencing

We prepared sequencing libraries using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® following the manufacturer's instructions and used 50 µl of samples (5-200 ng of DNA depending on specimen age and tissue size). Quality of libraries was assessed using a bioanalyzer to assure that library inserts were of the desired size and that adapters were removed before sequencing. The samples were sequenced by Novogene in San Francisco (USA) on an Illumina Novaseq 6000.

2.3.4. Bioinformatic analysis for phylogenetics

Analyses were conducted on a local scientific computing cluster at the California Academy of Sciences. Initially, 61 samples were sequenced and the program FastQC v.0.11.5 (Andrews, 2010) summarized sequence run data and quality for all the samples. For six individuals, the quality and depth was low and we eliminated them from further analysis. The analysis was continued with 55 individuals (Table 1), and the program FastP v.0.23.2 (Chen et al., 2018) was used to filter adapters and low-quality sequences. Subsequently, sequences were mapped to a high quality reference genome of *P. nanus* (Anchundia et al., 2024b, NCBI accession JAWZSU000000000) using the program BWA-MEM v.2.2.1 (Li, 2013). Resulting alignment files were kept in BAM format to reduce the file sizes. SAMtools v.1.15 (Danecek et al., 2021) flagstat was used to check basic statistics of the number of reads that passed map quality filters and how many did not adequately map to the reference. We used the program DamageProfiler v1.1 (Neukamm et al., 2021) on the 15 "ancient" DNA sequences to identify signatures of deamination. No sample showed signs of deamination, so the analyses continued with all samples.

The program Picard v.2.26.11 (Picard Tools Broad Institute) was used with two program functions, including MarkDuplicate to mark the repeated reads to avoid bias in the subsequent analysis and the AddOrReplaceReadGroups function to add read groups so that each sequence file had a unique ID and identified the sequencing platform. The program Freebayes v.0.9.2 (Garrison & Marth, 2012), a Bayesian genetic variant detector, was used to call variant sites and filter the variants among the 55 individuals analyzed. Single nucleotide polymorphisms (SNPs) and insertion/deletion sites (Indels) are labeled in the Freebayes output variant call file (VCF). For our analysis, we proceeded to work only with biallelic SNPs (18,802,482 SNP sites in our initial VCF) and the program VCFtools v.0.1.16 was used to split SNPs and Indels into separate files (Danecek et al., 2011). Additionally, two

samples (1 Santa Cruz and 1 Baltra) were removed at this step of the analysis because they had a significant number of sites with missing data, leaving 53 samples for the phylogenetic analyses.

VCFtools was used to filter SNPs using the following methods; map quality (MQ) > 30, read depth (minDP) > 4, meanDP > 4 and max meanDP = 50, missingness rate < 0.95, and minor allele frequency (MAF) > 0.05, obtaining an average coverage of 9.4x. After filtering, 233,916 SNPs remained in the dataset with 50 samples having between 0 and 8% missing data, 2 samples 13% and one sample 29% missing data. The filtering was increased with a missingness rate < 0.97 and 74,512 SNPs were obtained, where 52 samples had between 0 and 5% missing data and one sample had 17%. Finally, the maximum and strictest filtering was done with a missingness rate 1 and 13,523 SNPs were obtained where no sample had sites that were missing data. We continued the analysis with these three datasets (233,916 SNPs, 74,512 SNPs, 13,523 SNPs) to explore how robust the data was to various filtering thresholds. The vcf-to-tab utility (Chen, 2015) was used to convert the VCF data to tab-delimited text and vcf_tab_to_fasta_alignment.pl (Chen, 2015) was used to convert the data to FASTA format. In order to flatten and phase the VCF file on the heterozygous sites, IUPAC nucleotide ambiguity codes were substituted for SNPs for heterozygous individuals in the matrix using vcf-tab-to-fasta (Chen, 2015). Subsequently, the program ModelTest-NG (Darriba et al., 2020) determined the best-fit model to use for phylogeny. Of all the tested models, GTR-InvGamma with 4 gamma rate categories fit our data best. The seqmagick (FHCRC Computational Biology) program was used to convert the FASTA file to Nexus format for the Bayesian phylogenetic program MrBayes (Huelsenbeck & Ronquist, 2001). MrBayes was run with 4,000,000 generations with 2 runs and 4 chains. After the trees converged and reached a standard deviation less than 0.01, we discarded the first 25% of trees as burn-in, and performed a consensus analysis of the remaining sampled trees to estimate posterior probabilities of nodes and credible intervals of branch lengths. With the same data set, Maximum Likelihood tree and statistics were obtained using the RaxmlHPC, a high-performance computer program developed to handle large data sets (Stamatakis, 2006). We used the same GTR-InvGamma model with 50,000 rapid bootstrap replicates. With both phylogenetic methods, we ran several analyses with the different SNP datasets (233,916 SNPs, 74,512 SNPs, 13,523 SNPs) and obtained a similar topology and support values for each of the three datasets. Therefore, we used the trees of the data set with the greatest number of SNPs. Geneious Prime® software version 2023.0.1 (https://www.geneious.com) and FigTree 1.4.4 software (Rambaut, 2018) were used to visualize the phylogenetic trees and support values at nodes.

2.4. Structure analysis

The program STRUCTURE 2.3.4 (Pritchard et al., 2000; Hubisz et al., 2009) was used to perform Bayesian analysis of population structure among islands. This analysis was performed with the dataset of 13,523 SNPs. Only samples of *P. nanus* from islands with at least two individuals

sequenced were included in the analysis. A total of 49 individuals met this criterion. The STRUCTURE program was run with the data for 500,000 MCMC generations for each run, of which 50,000 generations were discarded as the burnin, and we used K values from one to eight with three runs for every K totaling 12,000,000 generations. We chose to run a maximum of eight K because we only had eight islands/populations in our analysis. Additionally, VCFtools was used to obtain population statistics including Weir and Cockerham FST values (Table 2).

2.5. Divergence time analysis

For the molecular clock analysis using biallelic SNPs, we used the SNAPP program (Bryant et al., 2012) and the dataset with 13,523 SNPs. Only one individual from each island was used. The XML file to run SNAPP was made using the Ruby script snapp_prep.rb (Matschiner, 2022), which added the list of individuals and constraints values. To calibrate the tree, we used the Galapagos *Pyrocephalus* population that is closest to the base of the phylogenetic tree (*P. dubius* from San Cristobal). The calibration value was obtained from the estimate from the analysis in Bayesian tree inference in BEAST for multi coalescent tree model, using a strict clock model. We used a nuclear DNA substitution average rate of 0.12% per million years, for passerines (Lerner et al., 2011). The SNAPP program by default applied ascertainment bias correction (Bryant et al., 2012), since the substitution rate is different in different parts of the genome. After obtaining the XML file, it was subsequently run as input in the BEAST program version 2.7.1 (Bouckaert et al., 2019) for 2 million generations with a burnin of 10% using the TreeAnnotator v2.6.3 program (Drummond & Rambaut, 2007). The Tracer program version 1.7.2 (Rambaut et al., 2018) was used to visualize the data and ensure that it converged with high confidence. Figtree version 1.4.4 (Rambaut, 2018) was used to visualize the tree along with its node support values and error estimates.

2.6. Bioacoustics of male song

Songs of adult males were recorded during the breeding season while visiting islands to collect blood samples for genetic analysis (Santa Cruz n=6, Pinzon n=4, Marchena n=4, Isabela n=4, Fernandina n=5, and Pinta n=2). Songs were opportunistically recorded while hiking and searching for birds. Each song file used in these analyses was assumed to be from different males because they were recorded in different geographic locations. Songs were recorded with a Zoom F1-SP field recorder equipped with a Zoom shotgun microphone capsule SGH-6 with a hairy windscreen to reduce wind noise and were saved in WAV format at high quality sample rate of 96 Khz-24 bit. One additional song from Santa Cruz was obtained from the avian bioacoustic repository xeno-canto (https://xeno-canto.org/) and was used in this analysis (audio file XC409621). Song spectrograms were plotted using the program Raven Pro 1.6 (Cornell Laboratory of Ornithology, Ithaca, NY, USA). From each territorial male recorded, we used the manual selection tool in Raven Pro to extract the minimum and maximum frequency in hertz (Hz) and the duration of each syllable (delta time in seconds (s)). These

measurements were then averaged for each individual and used in further statistical analyses. Due to the smaller sample size, the Pinta recordings were not used in these quantitative analyses but are still presented for qualitative comparison.

Song data were analyzed using R Studio (R Core Team 2022). We ran a Box-M test to test for equal covariance matrices across islands using the boxM function in the biotools package (Da Silva, 2021). Given homogeneity of the covariance matrices ($X^2 = 28.402$, df = 24, p = 0.24), we performed a linear discriminant function analysis (LDA) using the lda function from the MASS package (Venables & Ripley, 2002). We first scaled the predictor variables (low frequency, high frequency, and syllable length) to have a mean = 0 and sd = 1. Data were split into training and testing sets and then a LDA model was fit to the data. The LDA model was used to make predictions and a plot was created to visualize linear discriminants LD1 and LD2. Next, to test for differences in song between islands, a MANOVA was conducted with island as the independent variable and LD1 and LD2 (the canonical axes 1 and 2) as dependent variables. For each significant variable of the MANOVA, we conducted a univariate one-way anova using the kruskal_test function to determine which variables had significant differences between islands. For the resulting p-values, we applied a Bonferroni correction for multiple testing, so that $p \leq 0.0167$ was considered significant. For each significant variable, we then performed a Tukey HSD post hoc test for pairwise comparisons between islands, using the tukey_hsd function in the rstatix package (Kassambra, 2023).

2.7. Morphometric data

There is slight sexual size dimorphism in *P. nanu*s, with females being slightly smaller than males (Ridgway, 1897). Therefore, to avoid any gender bias in morphometric data, we only took measurements of fully adult male birds with red plumage to confirm the sex. We did not include morphometric data of female-plumed birds as both females from young males have similar plumage. We used wing length from live birds and bird skin collections for comparison between islands. We checked to ensure that primary feathers were intact to insure comparable wing measurements. Additionally, we included data from publications that included wing length measurements (Ridgway, 1897; Snodgrass & Heller, 1904), along with measurements we took when capturing birds for sampling. The data was grouped together and statistical analysis was conducted in R Studio (R Core Team, 2022). A Kruskal-Wallis test was run on wing length, and compared between islands. A pairwise Wilcoxon rank sum test was performed with Benjamini-Hochberg correction to determine which islands had differing wing lengths. The boxplot/figure was created in R.

2.8. Population genetics Santa Cruz

In order to more closely examine the population dynamics and genetic indicators of a declining population of vermilion flycatchers, we focused on the individuals from Santa Cruz. For this analysis, 14 individuals from Santa Cruz and one individual collected in Baltra were used (11 modern samples,

4 ancient samples) (Table 1). We include the specimen from Baltra since this is a satellite island of Santa Cruz separated by a distance of only 350 m (Fig. 1), and we assume that this represents the genetics of lowlands of Santa Cruz. Stringent filtering using VCFtools was applied to the biallelic SNPs, selecting only the SNPs with the following characteristics: missingness rate 1, minor allele frequency (MAF) > 0.05, mapping quality (MQ) > 30, read depth (minDP) > 4, meanDP > 4 and max meanDP 50, and Hardy Weinberg Equilibrium (HWE) > 0.05. As a result, no individuals had sites with missing data, giving 172,663 SNPs with an average depth of coverage of 11.09 ± SD 2.73 for ancient Santa Cruz samples and an average depth of coverage of 10.82 ± SD 2.3 for modern Santa Cruz samples (Supplementary material Fig. S1).

Using the program Plink (Purcell et al., 2007) with the Kinship function, we estimated the degree of consanguinity of modern individuals in Santa Cruz to look for signs of inbreeding (Fig. 6a). The VCFtools and Plink program were used with the "het" function to obtain information on heterozygosity from each individual and estimate the inbreeding coefficient, F (Fig. 6b). Additionally, VCFtools estimated population statistics including, Tajima D and nucleotide diversity (Pi π) with the window 100000 parameter to sample across the genome and estimate statistical values.

2.9. Population genetics across the Galapagos

From the islands where modern samples were collected (Isabela 13 samples, Pinzon 6 samples, Fernandina 4 samples and Pinta 4 samples), we estimated values of standard population statistics to infer the status of these populations. Each island population was considered a separate lineage and SNP variant data were separated into distinct files for each island. Stringent filtering using VCFtools was used on the biallelic SNPs, only SNPs with the following characteristics were selected: missingness rate 1, minor allele frequency (MAF) > 0.05, mapping quality (MQ) > 30, read depth (minDP) > 4, meanDP > 4, max meanDP 50, and Hardy Weinberg Equilibrium (HWE) > 0.05. The final data files included 1,465,429 SNPs for Fernandina Island, 1,002,638 SNPs for Isabela, 974,344 SNPs for Pinzon, and 836,158 SNPs for Pinta. The Plink program was used with the het function to obtain heterozygosity information from each individual and estimate the inbreeding coefficient F. VCFtools was used to obtain population statistics including, Tajima D and nucleotide diversity (Pi π) with the window 100000 parameter to sample across the genome and estimate statistical values.

3. Results

The average number of reads after sequencing for modern samples was 68,725,956 ± SE 2,966,047 and 93% of these reads passed quality control with a Phred score of Q > 30. Ancient samples were sequenced at greater depth due to small fragments sizes and the number of reads was

higher with an average of 103,645,890 ± SE 13,187,741, with 86% of these reads passing quality control with a Phred score of Q > 30.

3.1. Phylogenetics and divergence dating

Using Maximum Likelihood and Bayesian analyses for our data set with 53 samples with 233,916 SNPs, we recovered a well-supported phylogenetic tree (Fig. 2a). *Pyrocephalus obscurus* from Arizona USA and *P. dubius* from San Cristobal were used to root the tree. Additionally, by conducting a divergence time analysis for each island using the SNAPP program, we estimated the time of isolation from each clade (Fig. 3). With an estimated separation of *P. dubius* from the rest of the Galapagos archipelago between 1.16 to 1.40 MYA 95% highest posterior density (HPD). Within *P. nanus*, our tree supported at least six genetically isolated and monophyletic lineages (Fig. 2a). The Floreana lineage is sister to a clade containing all remaining *P. nanus*, suggesting that there is considerable genetic distance to its sister clade, with an estimated divergence date between 0.95 to 1.21 MYA 95% highest posterior distribution (HPD).

Within the remaining *P. nanus*, there are five distinct lineages that began to diverge between 0.68 to 0.86 MYA with a 95% HPD. The first split, around 0.77 MYA, separates a clade containing birds from Isabela and Fernandina Islands plus another lineage of birds from Pinzon and Rabida isolated between 0.41 to 0.55 MYA 95% HPD. Rabida birds are more closely related to Pinzon, despite Rabida's close geographic proximity to Santiago Island. Birds from Santa Cruz, along with birds from Santiago, Marchena, and Pinta islands, share a common ancestor and comprise the remaining clade. They diverged approximately 0.36 to 0.48 MYA 95% HPD, with an average estimate of Santa Cruz being isolated 0.42 MYA. Birds from Pinta Island are sister to Marchena and Santiago birds, but separated by a long branch, diverging between 0.22 to 0.31 MYA 95% HPD. Santiago and Marchena are a sister group separated between 0.1 to 0.19 MYA 95% HPD (Fig. 3).

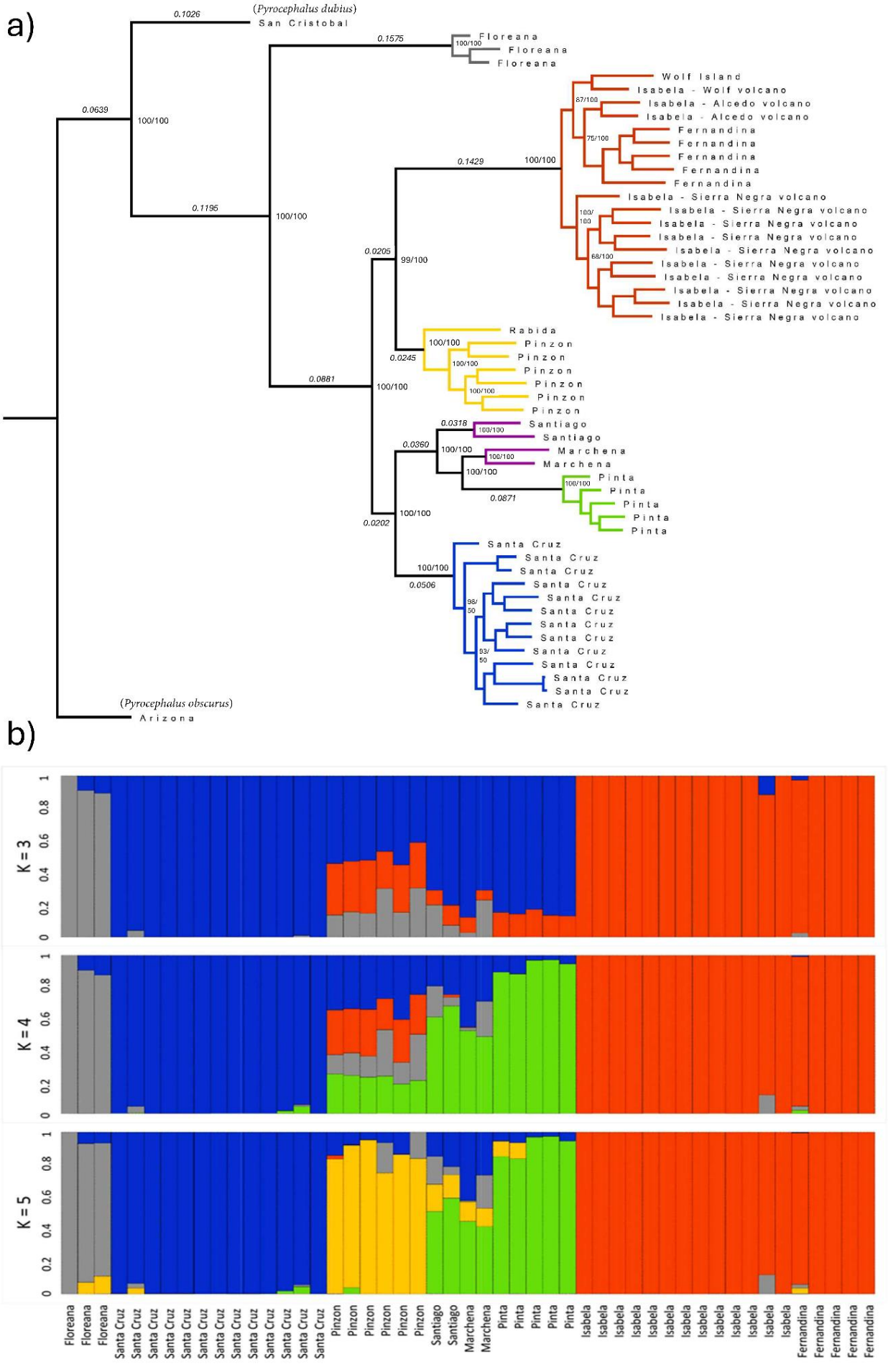

a)
(Pyrocephalus dubius)
San Cristobal
Floreana
Floreana
Floreana
Wolf Island
Isabela - Wolf volcano
Isabela - Alcedo volcano
Isabela - Alcedo volcano
Fernandina
Fernandina
Fernandina
Fernandina
Fernandina
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Isabela - Sierra Negra volcano
Rabida
Pinzon
Pinzon
Pinzon
Pinzon
Pinzon
Pinzon
Santiago
Santiago
Marchena
Marchena
Pinta
Pinta
Pinta
Pinta
Pinta
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
Santa Cruz
(Pyrocephalus obscurus)
Arizona
b)
K = 3
K = 4
K = 5
Floreana
Santa Cruz
Pinzon
Santiago
Marchena
Pinta
Isabela
Fernandina

Fig. 2. a) Phylogenetic ML tree display with 53 individuals. The support values reported at each node are ML bootstrap values / Bayesian posterior probabilities. Branch lengths are reported above branches in italics. b) STRUCTURE analysis of *P. nanus* populations among islands.

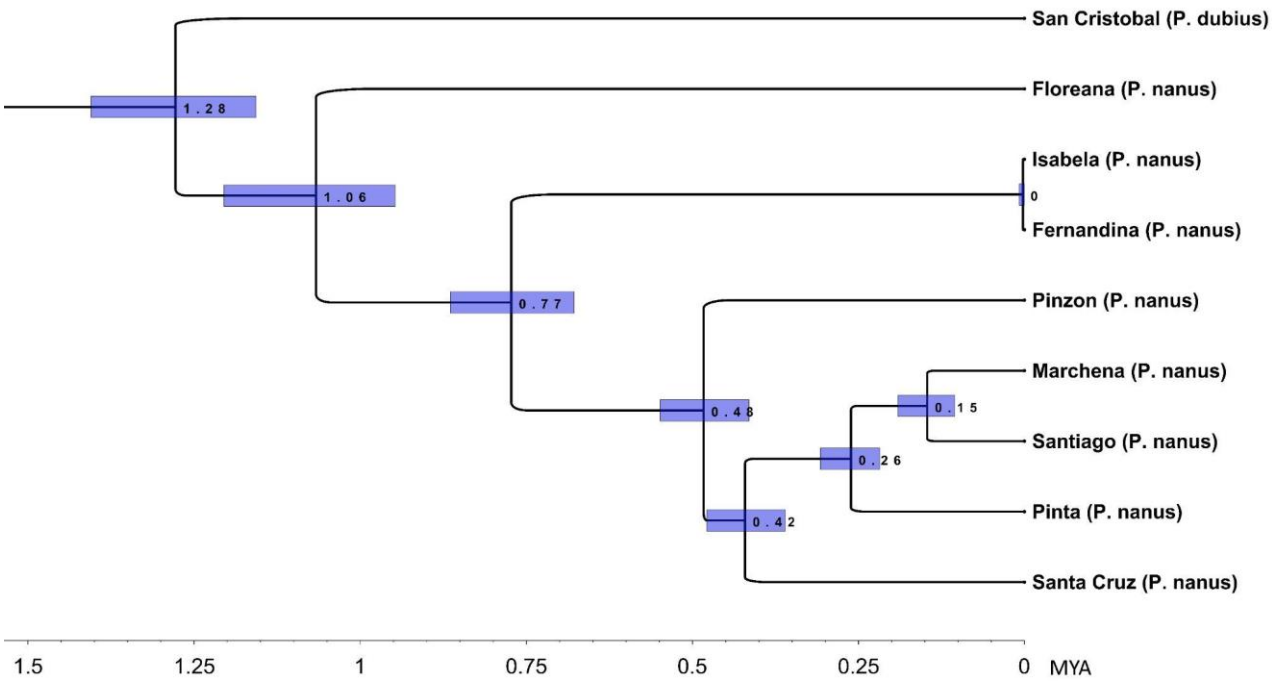


Fig. 3. Divergence times of Galapagos *Pyrocephalus* clades using SNAPP program for population of each island. The purple bars are the 95% probability that a population has been isolated and the number is the average in Mya of separation.

3.2. Island differentiation using STRUCTURE and FST

The island-wide population structure and differentiation analysis by the STRUCTURE program showed clearly the genetic division within various *P. nanus* subpopulations (Fig. 2b). The K = 3 shows at least three different major subpopulations, which include (1) Floreana, (2) Santa Cruz, Pinzon, Santiago, Marchena and Pinta, and (3) Isabela and Fernandina and up to K = 5, clearly recovers subdivision of the Pinzon and Pinta populations. This pattern coincides with the major clades in the phylogeny. Also, there is strong genetic variation observed in FST values (Table 2) between populations from different islands suggesting moderate and high levels of differentiation between these populations. Within *P. nanus*, the island pairs with the lowest FST value were Santiago-Marchena (FST=0.01) and Isabela-Fernandina (FST=0.02) and the islands with the highest FST value were Pinta-Floreana (FST=0.27). The *P. nanus* population from Pinta Island was the population with the highest degree of genetic differentiation from *P. dubius* from San Cristobal Island (FST=0.35) Supplementary material Fig.S2 (Table 2).

Table 2

Pairwise FST of *P. nanus* individuals from eight islands and *P. dubius* from San Cristobal Island.

	Pairwise Fst							
	Santa Cruz	Pinzon	Santiago	Marchena	Pinta	Isabela	Fernandina	Floreana
Santa Cruz	--							
Pinzon	0.13	--						
Santiago	0.14	0.12	--					
Marchena	0.14	0.14	0.01	--				
Pinta	0.14	0.16	0.15	0.18	--			
Isabela	0.16	0.12	0.14	0.16	0.15	--		
Fernandina	0.18	0.13	0.15	0.16	0.19	0.02	--	
Floreana	0.19	0.18	0.17	0.18	0.27	0.19	0.21	--
San Cristobal	0.25	0.26	0.21	0.22	0.35	0.22	0.24	0.26

3.3. Bioacoustic song analysis

Prior to the acoustic analysis, during sample collection on the Galapagos Islands, D.A. noticed that the males' territorial songs were different between islands. Acoustic analyses confirm that there are differences in male songs between different islands. The spectrogram shows that only the populations of Isabela and Fernandina islands maintain a similarity in the male song pattern with a frequency range of 3358 - 6661 Hz, while the male songs of Pinta Island have the largest frequency range of 358 - 14534 Hz. Santa Cruz has a frequency range of 2035 - 6752 Hz, Pinzon 1449 - 10464 Hz, and Marchena 1127 - 6911 Hz. The time duration of each syllable(s) was very similar for birds from Isabela, Fernandina, Santa Cruz, and Marchena with a duration of 0.13-0.14 s, while birds from Pinta and Pinzon had a time duration of each syllable of 0.16-0.17 s (Fig. 4a).

The Linear Discriminant Analysis (LDA) model distinctly separated songs by island (Fig. 4b) and all songs (100%) were classified correctly by island via cross-validation. LD1 explained 92.0% of the variation in the song data, with low frequency and high frequency as important variables in discriminating between islands. Songs were distinctly separated by island in LD1; however, Fernandina and Isabela were grouped close together. The LD2 accounted for 7.96% of the variation. Songs were significantly different between islands (Wilks lambda <0.001, $F_{4,26} = 307.6$, $p < 0.0001$). Songs significantly differed between islands in high frequency ($H(4) = 23.5$, $p < 0.001$), low frequency ($H(4) = 24.8$, $p < 0.0001$), and syllable length ($H(4) = 19.7$, $p < 0.001$), with a Bonferroni p-value cutoff of $p = 0.0167$. Pairwise comparisons between islands for each variable separately yielded significant differences between most islands in high frequency and syllable delta time (Table S1).

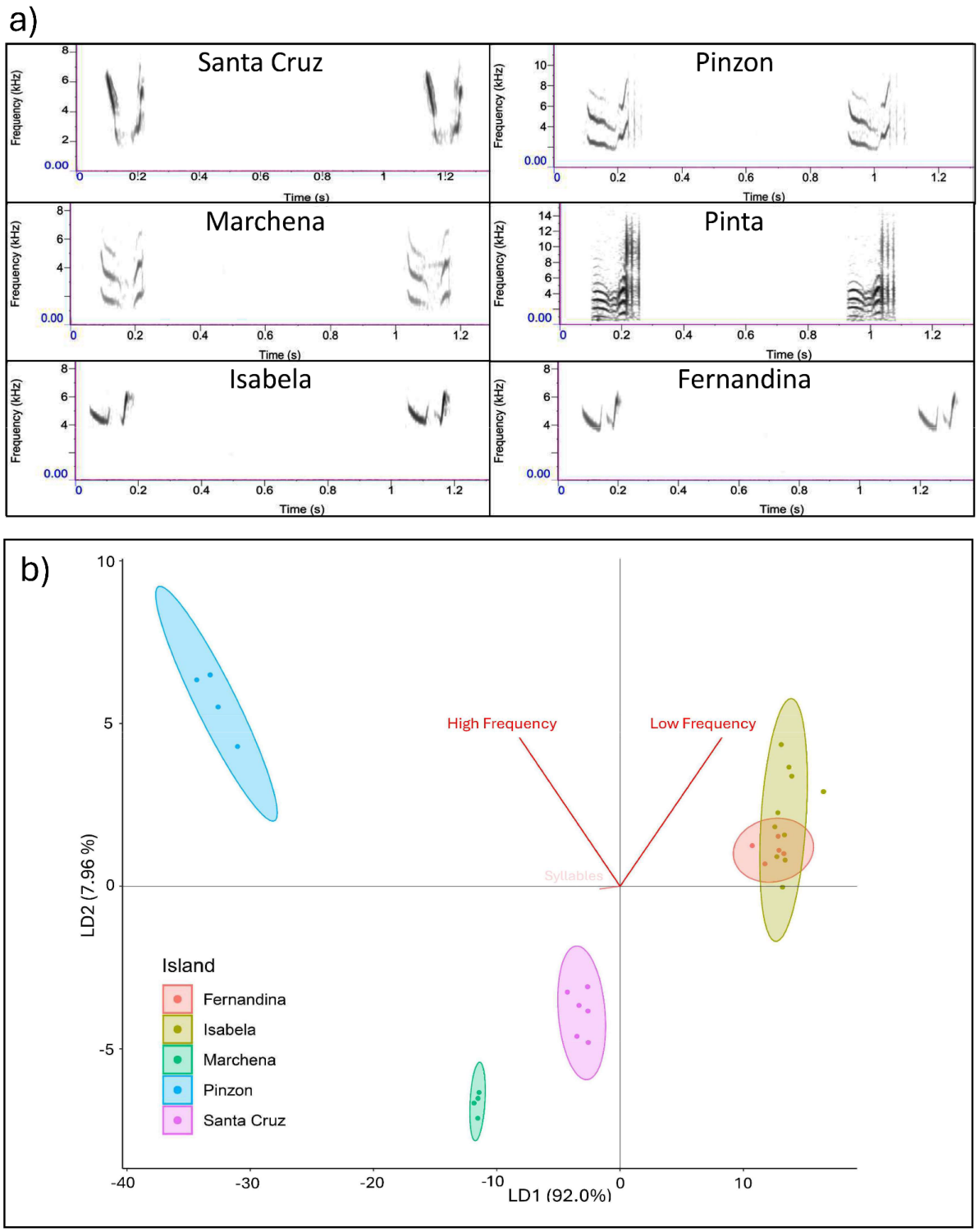


Fig. 4. a) Spectrogram comparison of male territorial and breeding/displaying songs during breeding season on six islands. There were no songs available from Floreana, San Cristobal, Rabida and Santiago. Each square is a spectrogram with two syllables with frequency (KHz) on the y-axis and time (s) on the x-axis. b) LD1 separates 4 groups due to variation in their song of birds from Pinzon, Marchena, Santa Cruz and a group consisting of Isabela and Fernandina. We see that the high frequency (Hz) and low frequency (Hz) variables are in a positive direction and these variables

account for most of the variation in the data, while the delta time of syllables (s) does not explain much variation.

3.4. Morphometrics - wing length

Wing length was significantly different between *P. nanus* males on some islands ($X^2 = 19.189$, df = 7, p = 0.0076). Pairwise comparisons indicate that both Pinzon and Santa Cruz have significantly different wing lengths compared to both Isabela and Pinta separately (Table 3, Fig. 5).

Table 3

Pairwise p-values for wing length between islands. Only males were included in the analysis. Significant p-values (<0.05) are in bold.

	Fernandina	Floreana	Isabela	Marchena	Pinta	Pinzon	Santa Cruz
Floreana	0.802	-	-	-	-	-	-
Isabela	0.370	0.617	-	-	-	-	-
Marchena	0.653	0.530	0.276	-	-	-	-
Pinta	0.191	0.372	0.617	0.269	-	-	-
Pinzon	0.370	0.191	**0.041**	0.639	**0.041**	-	-
Santa Cruz	0.642	0.370	**0.041**	0.967	**0.041**	0.372	-
Santiago	0.802	0.620	0.195	0.694	0.184	0.269	0.642

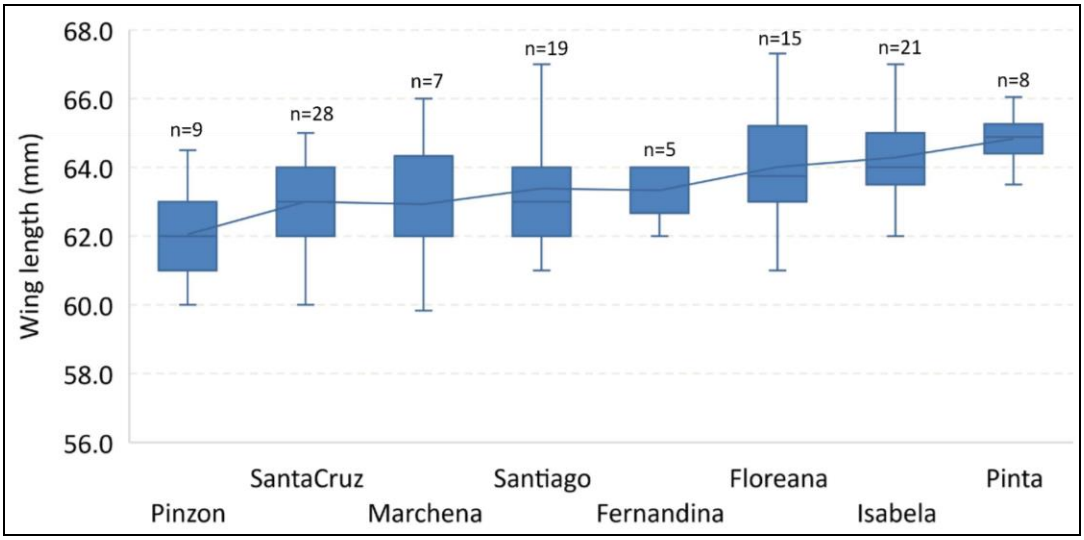


Fig. 5. Comparison of wing length of *P. nanus* populations from different islands. The top boundary of each boxplot reflects the third quartile, the bottom boundary the first quartile, and the lines at each end show the minimum and maximum value. The line inside the diagram is the mean.

3.5. Population genetics for Santa Cruz

3.5.1. Relatedness of modern samples from Santa Cruz

Although the current population of Santa Cruz is extremely small (30 to 40 individuals), we found that most of the 11 individuals we sampled randomly were not closely related genetically. From the 55 pair recombinations between these 11 individuals, only two showed a first and second degree of consanguinity. The rest of the recombination showed a low degree of relatedness. Additionally, eight birds (4 females and 4 males) that formed pairs during 2020 and 2021 were not related to each other (Fig. 6a).

3.5.2. Santa Cruz past and present inbreeding coefficient

The mean inbreeding coefficient of the modern population in Santa Cruz, using 345,326 alleles in using Plink program, was F = -0.026 ± SD 0.069 (range of -0.133 to 0.078), and using VCFtools program, was F = 0.007 SD ± 0.067 (range -0.095 to 0.108). While the average for past population using a similar number of alleles; Plink program, result F = 0.044 ± SD 0.247 (range of -0.254 to 0.324), VCFtools software F = 0.076 ± SD 0.238 (range -0.212 to 0.347) (Table 4, Fig. 6b). There is no obvious change in heterozygosity and thus no visible inbreeding. This agrees with the consanguinity analysis that the birds in the existing population are not related, therefore there are not cases of inbreeding.

Table 4

Inbreeding coefficient of birds from Santa Cruz Island from populations that lived in 1905, 1961, considered past or old samples (4 samples) and modern population where most of the samples were collected between 2019-2021, and only one sample in 2016 (11 samples). Hum (humid zone), trans (transition zone), E.HOM (expected homozygote), O.HOM (observed homozygote), and SNPs (single nucleotide polymorphism).

Year sample & habitat	Sex	No. SNPs	PLINK (O.HOM)	PLINK (E.HOM)	(F) inbreeding coefficient PLINK	VCFTOOL (O.HOM)	VCFTOOL (E.HOM)	(F) inbreeding coefficient VCFTOOL
1905 Arid past	M	172663	0.805	0.711	0.325	0.805	0.701	0.347
1905 Arid past	F	172663	0.752	0.711	0.141	0.752	0.701	0.170
1961 Hum/Trans past	F	172663	0.637	0.711	-0.254	0.637	0.701	-0.212
1961 Hum/Trans past	M	172663	0.701	0.711	-0.033	0.701	0.701	0.001
2016 Hum modern	M	172663	0.727	0.711	0.057	0.727	0.701	0.089

2019 Hum modern	M	172663	0.719	0.711	0.029	0.719	0.701	0.061
2019 Hum modern	M	172663	0.734	0.711	0.078	0.734	0.701	0.109
2019 Hum modern	F	172663	0.694	0.711	-0.058	0.694	0.701	-0.022
2019 Hum modern	M	172663	0.709	0.711	-0.007	0.709	0.701	0.026
2020 Hum modern	F	172663	0.699	0.711	-0.043	0.699	0.701	-0.008
2020 Hum modern	F	172663	0.677	0.711	-0.117	0.677	0.701	-0.080
2020 Hum modern	F	172663	0.683	0.711	-0.097	0.683	0.701	-0.061
2020 Hum modern	M	172663	0.711	0.711	0.001	0.711	0.701	0.034
2020 Hum modern	F	172663	0.672	0.711	-0.133	0.672	0.701	-0.096
2021 Hum modern	M	172663	0.711	0.711	-0.001	0.711	0.701	0.032

3.5.3. Nucleotide diversity and Tajima's D

There was a significant reduction in nucleotide diversity ($\boldsymbol{\pi}$) (22.5 %) from past to modern samples from Santa Cruz ($p < 0.001$), with a mean $\boldsymbol{\pi}$ for past samples of 0.40 ± SE 0.0003 and a mean $\boldsymbol{\pi}$ for modern samples of 0.31 ± SE 0.0003 (Supplementary material Fig.S3a). Also, there is a significance difference in Tajima's D between modern and ancient samples ($p < 0.001$). The Tajima's D values estimated from ancient samples show the Santa Cruz population did not deviate significantly from the expectation of a neutrally evolving population with constant size, with Tajima D = 0.14 ± SD 0.60. In contrast, the modern population has a Tajima's D value of 0.41 ± SD 0.68, so the population experienced a decrease in population size, indicated by the low levels of low and high polymorphism (Supplementary material Fig.3b).

3.6. Population genetics of Fernandina, Isabela, Pinta, and Pinzon

Of the four islands with current populations sampled apart from Santa Cruz, Pinta (F = -0.15 ± SE 0.03) and Fernandina (F = -0.14 ± SE 0.02) had the lowest inbreeding coefficient, followed by Pinzon (F = -0.11 ± SE 0.06). The birds with the highest inbreeding coefficient were in Isabela (F = -0.02 ± SE 0.02). This coefficient coincides with the order of the average coefficient of nucleotide diversity in these four islands. Firstly, Pinta has a higher nucleotide diversity ($\boldsymbol{\pi}$) (0.40 ± SD 0.12), followed by Fernandina (0.40 ± SD 0.12), Pinzon (0.36 ± SD 0.13), and Isabela, which had the lowest ($\boldsymbol{\pi}$) coefficient (0.33 ± SD 0.14). Additionally, Pinta was the island with the lowest Tajima D value (0.07 ± SE 0.008), followed by Fernandina (0.21 ± SE 0.004), these populations did not have a significant deviation from the neutral expectation. These Tajima D values were followed by Pinzon (0.41 ± SE 0.007) and Isabela with the highest Tajima D value (0.87 ± SE 0.005), which indicates these two island populations might have experienced decreases in population size recently.

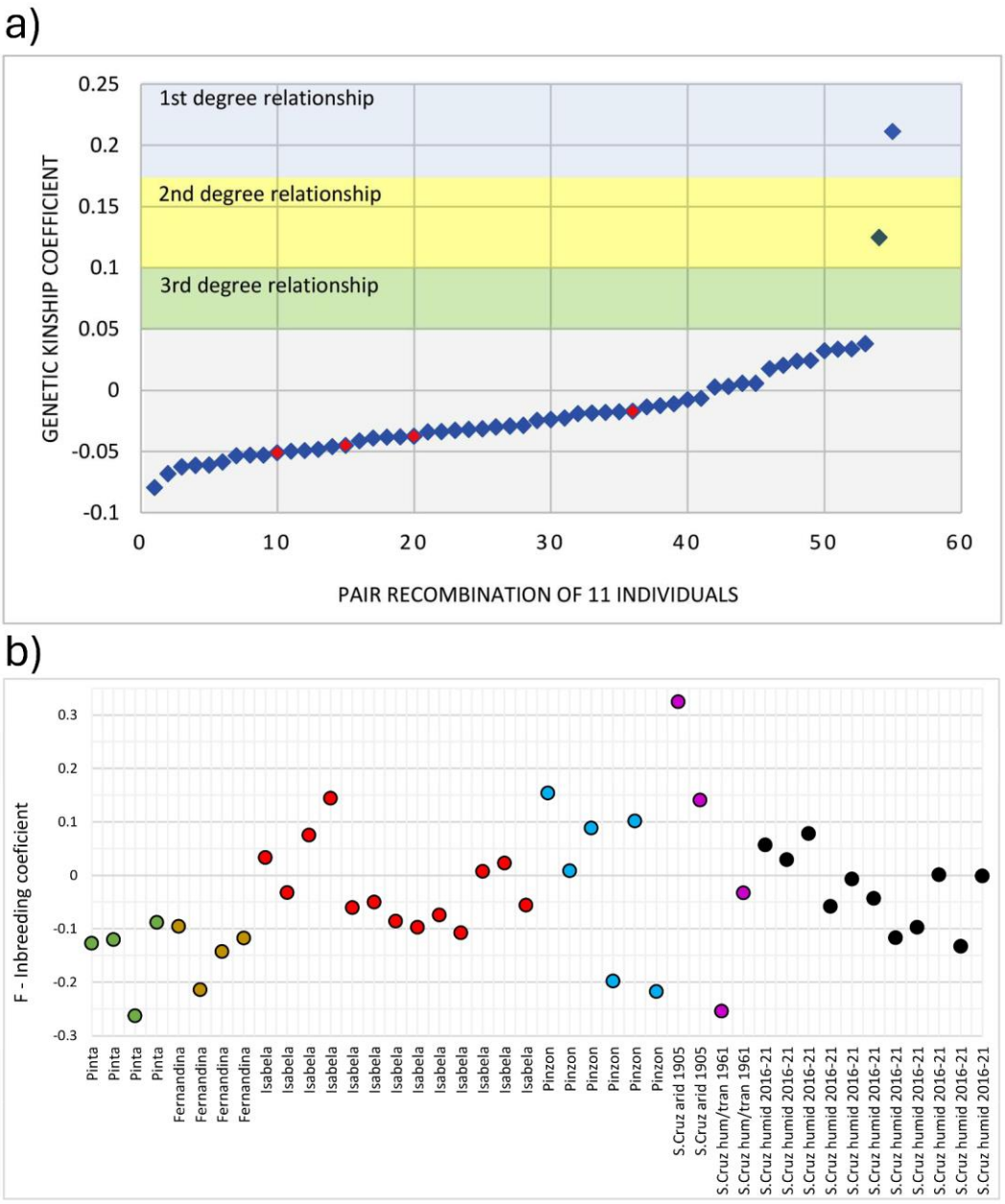


Fig. 6. a) Consanguinity of 11 sequenced samples from individuals in Santa Cruz in 55 pairwise comparisons. 1st degree relationships are very closely related individuals, one male and one female (parent-offspring, full sibling). 2nd degree relationships are two closely related females (half siblings, grandparent-grandchild, etc.). Known breeding pairs (shown in red) were not related. b) Inbreeding coefficients from modern samples collected from Pinta (green dots), Fernandina (brown dots), Isabela (red dots), Pinzon (blue dots), Santa Cruz from 1905-1961 (purple dots), and Santa Cruz from 2016-21 (black dots).

4. Discussion

Populations of *P. nanus* are highly genetically differentiated among the Galapagos islands. The acoustic, morphological, and large genomic dataset, using hundreds of thousands of SNPs, clearly indicates that this is an older and more taxonomically complex group than previously thought. These results provide a better understanding that should ultimately improve the taxonomy as well as management options for Galapagos *Pyrocephalus*. The isolation and differentiation of at least six groups were detected in the phylogeny and the STRUCTURE analysis: 1) Floreana, 2) Isabela+Fernandina, 3) Pinzon+Rabida, 4) Santiago+Marchena, 5) Pinta, and 6) Santa Cruz. Our results appear biogeographically consistent with these islands' ages, geographic locations, and relative isolation. In addition to genetic differentiation, we also present data on significant differences in male breeding songs between most populations and significant differences in wing morphometrics between some populations. Likewise, Swarth (1931) mentioned that there are differences in *P. nanus* between islands. Our results also confirm that *P. dubius* from San Cristobal belongs to a separate species. This had already been suggested previously (Carmi et al., 2016) and accepted (Chesser et al., 2022). This species coalesces with *P. nanus* around ~1.3 MYA.

4.1. Differences of the populations between islands

4.1.1. Floreana

Surprisingly, in our analyses, the *P. nanus* individuals from Floreana form a separate isolated clade. The separation of Floreana with the other population of *P. nanus* clade is deep, which suggests that Floreana population has been in isolation for a very long time, with a mean divergence time of 1.06 MYA. Floreana Island is one of the older and more isolated islands in the Galapagos archipelago that has developed several of its own endemic birds including the Floreana Mockingbird (*Mimus trifasciatus*) (Arbogast et al., 2006), Medium Tree Finch (*Camarhynchus pauper*) (Kleindorfer et al., 2014), endemic subspecies of Grey Warbler Finch (*Certhidia fusca ridgwayi)* (Grant et al., 2005), genetically isolated populations of yellow warblers (*Setophaga petechia*) (Chaves et al., 2012), Galapagos flycatchers (*Myiarchus magnirostris*) (Sari and Parker 2012) and a genetically differentiated Short-eared Owl population (*Asio flammeus galapagoensis*) (Schulwitz et al., 2018).

Similarly, the morphological differences found by Ridgway (1897) led him to name the vermilion flycatcher population a different species, *Pyrocephalus carolensis*. Our phylogenetic and STRUCTURE analyses support the differentiation of the Floreana population. Similar observations were made in the Pinchot Expedition to the Galapagos Islands in 1929, suggesting that *P. nanus* "females from Floreana compared to Santa Cruz are quite different and warrant a taxonomic revision

to at least subspecies category" (Fisher & Wetmore, 1931). Swarth reached a similar conclusion in 1931 when he grouped birds from Santa Cruz and compared them with other islands, mainly Floreana, and concluded that the populations between islands are quite different and the differences are most noticeable in females (Swarth, 1931). In particular, Floreana females have a very marked chest with streaks, while in Santa Cruz they do not. During our inspection of the specimens preserved in the CAS museum in 2021, we confirmed this pattern. The Floreana population was formerly very large and present throughout the island (Rothschild & Hartert, 1899; Baldridge, 1968; Steadman, 1986). Unfortunately, this population appears to have already gone extinct from Floreana Island (Merlen 2013, Dvorak et al., 2021) and intensive bird surveys in 2008 (O'Connor et al., 2010), yearly from 2014 to 2016 (Dvorak et al., 2017), and 2023 survey (CDF unpublished data) could not find any *P. nanus* individuals. The last verified observations of *Pyrocephalus* on Floreana were in 2004 (O'Connor et al., 2010). It is possible that the decline of this population began in the 1970s, coinciding with the arrival of the Avian Vampire Fly, *Philornis downsi*, as its earliest record in Galapagos is from 1964 (Causton et al., 2006). This avian parasite is one of the main threats to flycatchers species in Galapagos (Leuba et al, 2020; Pike et al., 2021; Mosquera et al., 2022), as it drastically reduces breeding success and recruitment of new individuals into the population. Additionally, another major threat is the alteration of habitat and food sources, which produces a decline in their reproductive success (Anchundia et al., 2024a).

The remaining populations of *P. nanus* form a sister clade to Floreana. Within this clade, there are two large separated groups, with a mean divergence time of 0.77 MYA. One clade includes an Isabela-Fernandina group and a Rabida-Pinzon group and the second clade includes a Santa Cruz group, Pinta group and Santiago-Marchena group.

4.1.2. Isabela and Fernandina

These two islands form a separate isolated group with a very deep branch, which separated around 0.77 MYA. Due to the geographical proximity between Isabela and Fernandina (~4 km), as well as the younger age of these islands (particularly Fernandina), the phylogenetic and STRUCTURE analyses suggest the movement of birds between these two islands and probable colonization of Fernandina from Isabela and a high level of genetic exchange between Isabela's volcanoes. We also observed that the populations of both islands have similar male songs during the nesting season, which is consistent with genetic similarities of both islands.

We found no evidence of recent genetic exchange between this group and the other groups of *P. nanus*. This is somewhat surprising since the other groups of *P. nanus* are only 16 km away from

Santiago Island and 18 km away from Pinzon Island (Fig. 1). Despite the fact that this distance seems close for a highly mobile bird, our data suggest that there is little to no genetically-effective migration. This is striking since a bird was collected from Wolf Island (Gifford, 1919), which is 143 km from Isabela Island and the phylogenetic analysis placed this bird within this Isabela-Fernandina group. This shows that they can disperse over longer distances, but perhaps there is selection against these migrants on Pinzon or Santiago where these populations are genetically divergent from the Isabela-Fernandina populations.

In the population genetic analysis of Isabela, most of the samples were from the Sierra Negra volcano. The results showed less genetic diversity compared to other modern populations on Fernandina, Pinzon, and Pinta. This result coincides with the population reduction observed in the Sierra Negra volcano of Isabela Island observed by local people (CDF unpublished data). On the contrary, the Fernandina population is the modern population with the highest genetic diversity, but we lacked studies on this population due to the remoteness of this active volcano. Most of Fernandina Island is inhospitable, with large barren lava fields; however, the LVF population is thriving in scattered patches of vegetation.

4.1.3. Pinzon and Rabida

Another lineage discovered includes the birds from Pinzon and Rabida, which have been separated for approximately 0.48 MYA. The geographically closest island to Pinzon is Santa Cruz, separated by a 10.5 km ocean channel. However, we find that Pinzon's population has remained isolated, developing a unique type of song. Surprisingly, its wing size was also found to be the smallest of all *P. nanus* populations in Galapagos. It is likely that the smaller size of their wings is an evolutionary adaptation to the type of ecosystem where they live. Pinzon Island has dense xerophytic vegetation, consisting mainly of bushes. Perhaps, the smaller wing size would make it easier for the birds to fly efficiently between close vegetation to chase and hunt insects (Anchundia, pers. obs. on Pinzon Feb 2023; Gerdes et al., 2012). It also could help them to hover and capture insects on the branches (Anchundia, pers. obs on Pinzon Feb 2023; Warrick, 1998). Additionally, the population genetic analysis of Pinzon shows a slightly lower genetic diversity than other islands. This coincides with the observations made one decade ago that this population was reduced due to the impact of introduced black rats (*Rattus rattus*) (Island Conservation, unpublished data). However, currently its population has increased (CDF, unpublished data) due to rodent eradication conducted on this island in 2012 (Rueda et al., 2019).

4.1.4. Santa Cruz

In the second largest clade of *P. nanus,* there is greater complexity. The results showed that the Santa Cruz population diverged from Santiago, Marchena and Pinta approximately 0.42 MYA. This population developed its own song, and after the Pinzon population it is the population with the smallest wing size. Swarth (1931), with morphological analysis, separated the population of Santa Cruz into an endemic subspecies, *P. nanus intercedens.* Similarly, Ridgway delineated Santa Cruz as a different species, *P. intercedens*, together with Isabela. These statements show that there are some morphological differences in the Santa Cruz population, compared to other islands. Our phylogenetic and STRUCTURE analyses validate the interpretation that this population has been isolated for a long time and forms a unique lineage. Our modern observations of live birds of plumage of females in this population compared with Pinzon, Pinta, Isabela and Fernandina show subtle differences. The females in Santa Cruz have been observed having a creamy yellow plumage on the abdomen up to the neck, and sporadically, females with orange plumage on the head have been recorded. Unfortunately, this unique lineage is on the brink of disappearing due to its drastic population decline (Anchundia et al., 2024a).

4.1.5. Pinta

Geographically, this island is one of the most remote locations where *P. nanus* has established a population. Our results indicate that this population has been highly isolated, with a mean divergence time of 0.26 million years ago (MYA) and is rapidly differentiating from other populations. This population has also differentiated in its song from other islands and is the population with the largest wing size of all *P. nanus* populations, with a significantly larger size compared to Santa Cruz and Pinzon. Head-bill length was measured for four modern individuals from Pinta and their head-bill length was larger than other populations; however, we had insufficient data to conduct a more in-depth analysis. Ridgway (1897) divided this population into *P. abingdoni* together with Marchena. Similarly, Snodgrass and Heller (1904) considered these populations as the subspecies *P. nanus abingdoni*. Our phylogenetic and STRUCTURE analyses validate that there is a separation and isolation of this island from nearby Marchena Island. Due to the remoteness of Pinta Island, more information on its population in the past is lacking. However, the population genetic analysis shows that Pinta's population has a greater genetic diversity than other islands, which suggests that this population has remained stable. Pinta is the smallest island (60 km^2) in Galapagos with a humid zone and has diverse types of ecosystems with a steep elevation up to the summit volcano. The surprisingly larger wing size perhaps is an adaptation for frequent movement between regions of these ecosystems, since this bird is observed from the beach to the top of the highland zone of this island, 659 m above

sea level. Moreover, larger sized wings may also give the birds the ability to return to Pinta if they are blown away from the island by wind.

4.1.6. Santiago

It was not possible to obtain new samples from Santiago because *P. nanus* individuals have become scarce there. However, the two historical samples used in the analysis show that this population also remained isolated. Santiago is one of the largest islands in the Galapagos, with a variety of ecosystems. This island was one of the main islands visited by pirates, whalers, and explorers before Galapagos was inhabited. Records of these previous visits show that *P. nanus* was abundant from the coast to the highlands of Santiago (Colnett, 1798; Porter, 1822). Additionally, Darwin visited this island in 1835 and collected several specimens and the origin of the name *nanus* comes from this population (Gould, 1841, Ridgway, 1897). However, this unique lineage of Santiago is about to disappear. Therefore, more effort should be devoted to search efforts, update its population status, and its conservation. The last sighting was made in 2017 by park rangers in the highlands of Santiago. During our visit in 2020 to collect samples for this study, no birds were seen despite hiking 30 km. The disappearance of *P. nanus* might be linked to the destruction of nearly all the vegetation by introduced goats (Hamann, 1981; Schofield, 1989). The goats were introduced in 1813 by the U.S.S. Essex (Porter, 1822) and their population expanded dramatically to 100,000. They were eradicated on this island in 2005 (Cruz et al., 2009), but the *P. nanus* population did not recover. After the goat eradication, the invasive plant species blackberry expanded rapidly since there were no large herbivores to control it (Renteria, et al., 2012). Also, blackberry covers previously open areas, which leads to reduced foraging success of *P. nanus* (Anchundia, et al., 2024a). Furthermore, the Avian Vampire Fly is present in relatively high numbers, at least in the highlands of Santiago (CDF, unpublished data), adding further pressure to this population and perhaps reducing its reproductive success.

4.1.7. Marchena

The analysis also shows the isolation of this population, and the genetically closest population seems to be birds of Santiago Island, as the mean divergence time between the two islands is only 0.15 MYA. Geographically, Marchena is closer to Pinta, but there is no evidence of recent genetic exchange and our results indicate they separated ~0.26 MYA. Also, Marchena and Pinta have a high pairwise FST, suggesting that there has been no exchange between these islands for some time. However, Marchena and Santiago have a lower pairwise FST, suggesting a more recent connection. We do not know if the territorial song is unique to this island since we do not have songs from

Santiago Island, and both populations are more closely related genetically. Little is known about this isolated and remote population. It lives only in an arid and dry ecosystem. Reports from early explorers suggest the bird was not common (Rothschild & Hartert, 1899), hence few individuals were collected in the past. Most of the island (2/3) is covered with lava fields and the little vegetation present is in rocky places with little soil, so it does not have a proper environment to support large populations.

4.2. Changes within populations and speciation

One barrier to gene flow are the differences in song between bird populations since it is an important aspect of species recognition and mate choice (Grant & Grant, 2008). Differences in songs have served as a tool to delimit and classify species in avian systematics (Alstrom & Ranft, 2003; Kumar, 2003). For morphologically very similar bird species, acoustic and genetic differences have provided better resolution for their classification (Päckert et al., 2004; Toews & Irwin, 2008; Carpenter et al., 2022). Within *P. nanus*, it is evident that there are some acoustic differences between males in territorial or displaying songs.

Another diverse group in Galapagos similar to *Pyrocephalus* are mockingbirds (MB) (genus *Mimus*) (Arbogast et al., 2006). Phylogenetic analysis of this genus showed a similar distribution pattern to *Pyrocephalus*. The San Cristobal MB became isolated and split off, forming an endemic species. Similarly, the Floreana and Española MBs also split separately and became two endemic species (Arbogast et al., 2006; Hoeck et al., 2010; Nietlisbach et al., 2013). In the western part of the archipelago, MBs from Isabela, Fernandina and even the remote Wolf Island group in the same clade (Nietlisbach et al., 2013; Arbogast et al., 2006) with surprisingly low gene flow to the central islands, similar to *P. nanus*. MBs in the center of the archipelago were grouped together with Santa Cruz, Santiago, Santa Fe, Marchena and Pinta (Nietlisbach et al., 2013; Arbogast et al., 2006). This similar distribution pattern of the genus *Mimus* in Galapagos with *Pyrocephalus* is striking, suggesting that the same barriers impact these two groups. The *Mimus* group arrived in Galapagos with a mean divergence time of 0.5 MYA (Nietlisbach et al., 2013). It is a much younger group than *Pyocephalus*, which is estimated to have arrived in Galapagos around 1.3 MYA. However, *Mimus* populations in Galapagos have already been divided into four endemic species and six subspecies. Additionally, phylogenetic reconstruction using millions of SNPs from Darwin's finches, shows that Darwin's finches diverged around 0.9 MYA (Lamichhaney et al., 2015) and they have become 17 recognized species. There was a rapid radiation of ground and tree finches around 0.3 to 0.1 MYA (Lamichhaney et al., 2015). Although *Pyrocephalus* in Galapagos is as old as other groups of iconic and endemic Galapagos birds, its taxonomy has not been evaluated in detail.

Strong genetic differentiation, a lack of evidence of migration, song differentiation between populations and morphological differences strongly suggests there are several unique lineages or species within *P. nanus* (Fig. 7). Therefore, we must highlight the protection of these unique, genetically distinct populations. We suggest dividing *P. nanus* into at least six species. The Floreana population could be elevated and recognized as *Pyrocephalus carolensis* by Ridgway 1897; however, it is unfortunately extinct. The Isabela-Fernandina populations (*Pyrocephalus albemarlei*), Albemarle is the old name from this island, Santa Cruz (*Pyrocephalus intercedens*) by Swarth 1931, Pinzon (*Pyrocephalus duncan*) Duncan is the old name of this island, Pinta (*Pyrocephalus abingdoni)* by Ridgway 1897 and Santiago-Marchena (*Pyrocephalus nanus*) by Gould 1839. These groups are unique as they have been completely isolated (Fig. 7).

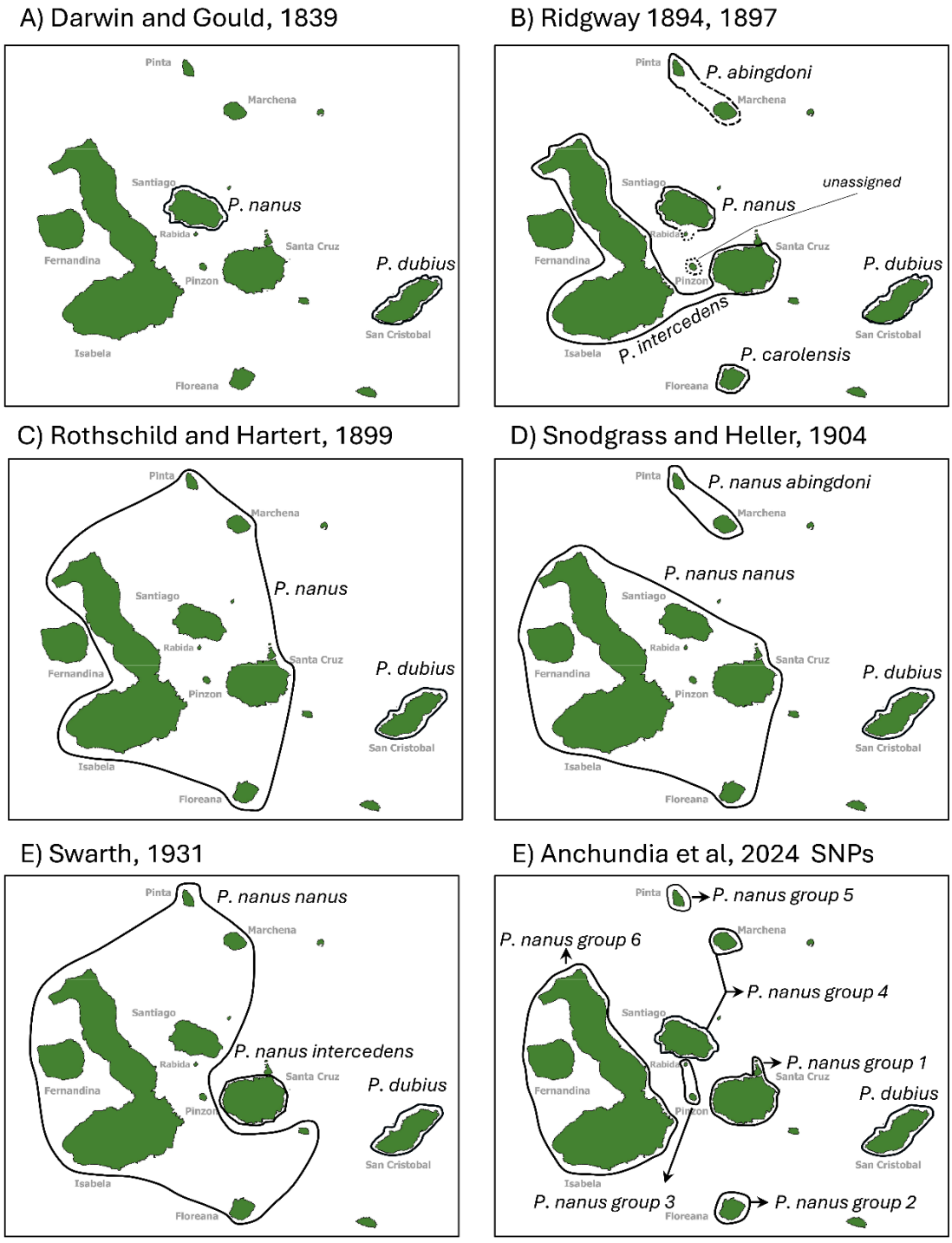


Fig. 7. History of the taxonomic classification of the genus *Pyrocephalus* in the Galapagos archipelago, it shows the references for old names used. Reference taken from the study by Carmi & Dumbacher 2016, dotted lines are islands where there is uncertainty if that island should be included by the author.

4.3. Santa Cruz population genetics

The Santa Cruz population (a unique island evolutionary unit) shows that it has lost genetic diversity and has lost some of its heterozygosity, although no strong signs of inbreeding are observed yet. Perhaps this is due to the rapid decline of the population, with low breeding success and the few individuals remaining are not closely related. However, inbreeding would not inhibit a population from recovering and growing again (Johnson et al., 2011). The genetic samples of birds used for the modern genetic analysis were contemporaneous and lived at the same time (years 2019-2021), except for one sample from 2016. This gives a good snapshot in time of this population. However, a visit to this small population in 2023-2024 showed that most of the birds sampled and used in this study were no longer present and may have died, as the average lifespan of these birds is short around 5- 6 years.

To prevent the disappearance of this population, a habitat restoration coupled with control of parasitism by the Avian Vampire Fly was implemented (Anchundia et al., 2024a) and since 2021, the reproductive success of these birds has improved. In 2021, eight fledglings from three pairs were incorporated into the population (Anchundia et al., 2024a). In 2022, seven fledglings from three breeding pairs, in 2023, 12 fledglings from three breeding pairs and in 2024, 15 fledglings from seven breeding pairs (CDF unpublished data) were incorporated into the population. Despite intensive search efforts, no successful nests from other pairs outside the managed area were recorded. In line with the increasing breeding success the cases of inbreeding between these new birds might increase too as they come from very few breeding pairs and many of the new fledglings are related. Thus, continued genetic evaluation of this vulnerable population is necessary.

The analysis of birds collected in 1905 shows a surprisingly high inbreeding coefficient. One of these samples is from Baltra, where its population size apparently was small during the capture, since Gifford (1919) mentioned only two individuals were observed. The other sample was from the arid zone on the northwest side of Santa Cruz. Deamination could occur throughout time specially in ancient samples, which could slightly alter the analysis. However, in our analysis, we did not find a problem with deamination.

4.4. Genetic rescue Santa Cruz

The Santa Cruz population is very small and potentially very closely related since the more recent higher breeding success from 2021 to the present. A genetic rescue through translocation has been suggested to prevent this population from declining further and disappearing. Genetic rescue is being used to avoid the extinction of populations that are declining, to prevent or limit inbreeding, and to alleviate deleterious effects and genetic drift (Johnson et al., 2011; Bateson et al., 2014; Segelbacher

et al., 2022; Jackson et al., 2022). However, an ecological, genetic, and risk assessment must be carried out on a target population for translocation before moving individuals (Frankham et al., 2017, Pérez-Pereira et al., 2022).

Genetic adaptations that may have evolved in a population in isolation could be altered by introducing a population with different adaptations, though there could be problems of introgression or outbreeding depression (Frankham et al., 2017; Rick et al., 2019, Pérez-Pereira et al., 2022). This could cause a reduction in fitness due to the mixture of two different genetic groups (Rick et al., 2019; Pérez-Pereira et al., 2022). Also, genes from one population could dominate genes from another population and could reduce genetic diversity (Frankham et al., 2017) and it might produce the opposite effect and accelerate its extinction. Moreover, for a successful genetic rescue of *P. nanus* on Santa Cruz, there must be a simultaneous effort to address the causes of the problems why the population became small, or else it will fall back into inbreeding after the translocation (Hedrick et al., 2019).

The genetic analysis shows that the Santa Cruz population is genetically distant from other islands. Therefore, any translocation to Santa Cruz could dilute its the genetic makeup. Thus, the best is to safeguard this genetically unique Santa Cruz population given that we know what is needed, such as a combination of habitat restoration and parasite control. But if the population continues to decline and inbreeding is severely affected, there would be no other option than to translocate individuals from the island with a closer FST, such as Pinzon. This is also the geographically closest population to Santa Cruz, separated by only 10.5 km. Also, the adaptation of a bird before translocating it to a similar type of ecosystem should be reviewed to understand its adaptation to this new environment.

Review of genetic adaptations of *P. nanus* living in the humid and arid ecosystems on Santa Cruz and Isabela shows no differentiation within an island's FST, which suggests there is movement of birds from humid and arid zones on the islands. Therefore, birds from humid or arid zones should be able to adapt to any of the two environments. The birds that live in Pinzon do so in an arid and xerophytic ecosystem, very similar to more than half of the surface area of Santa Cruz. However, the last remnant of the population in Santa Cruz is found in the humid zone (Anchundia et al., 2024a). Thus, if a translocation ever occurs, Pinzon birds moved to Santa Cruz would eventually need to adapt to a more humid environment.

4.5. Reintroduction of *Pyrocephalus* to Floreana

Currently there is a project underway to eradicate invasive species (cats and rodents) on Floreana Island (Ruíz-Ballesteros & Tejedor, 2022). Once the eradication is completed, there is

another project planned to reintroduce species that were present on this island and that previously disappeared due to anthropogenic activities (Hanson & Campbell, 2013), including *Pyrocephalus.* The ideal aim is to bring a population of *Pyrocephalus* from another island close to the Floreana lineage, to better adapt to this island ecosystem. Two candidate populations may be birds from Pinzon Island and birds from Isabela. Both populations have a FST of 0.18 - 0.19 respectively compared to the endemic Floreana populations, and both populations have a large population size (CDF unpublished data), which is important since a large population will be a constant source to establish a new population on another island.

Conclusions

1.- *Pyrocephalus* is a quite old group in Galapagos, even older than several other endemic bird groups in Galapagos.

2.- The geographical barriers of each island produces isolation of *Pyrocephalus*. Even between very close islands there is no movement.

3. - For the great time of isolation of some populations, accumulation of mutations, differences in morphology and song we suggest dividing *Pyrocephalus nanus* into 6 subspecies or species.

4.-The current population of Santa Cruz has decreased in genetic diversity compared to the past, although there were no strong signs of inbreeding. The current management of the remaining population needs to be continued and expanded and the genetic evaluation added to the management plan.

5.- After removing the problems that were the reasons for the disappearance of the Floreana birds, a reintroduction can be considered. The birds that would probably adapt best would be populations from Isabela, namely from the Alcedo and Sierra Negra volcanoes.

6.- The conservation status of each population should be evaluated on all the islands. To evaluate conservation strategies and highlight the protection of each lineage.

7.- If the taxonomy of *Pyrocephalus* in Galapagos is modified it would be necessary to use names that were already used in the past to describe species or subspecies.

CRediT authorship contribution statement

David Anchundia: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – original Draft, Writing – review & editing,

Visualization, Project administration, Funding acquisition. **Athena Lam:** Methodology, Investigation, Resources, Writing – Review & Editing, Supervision. **James B. Henderson:** Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Review & Editing. **Matthew Van Dam:** Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Review & Editing, Visualization. **Courtney Pike:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Review & Editing, Visualization. **Sabine Tebbich:** Writing – Review & Editing, Supervision, Funding acquisition. **Birgit Fessl:** Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition. **Heinz Richner:** Conceptualization, Methodology, Writing – Review & Editing, Supervision. **John Dumbacher:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing. Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was conducted with the permission of the Galapagos National Park Directory research permits PC-30-19, PC-11-20, PC-06-21, PC-22-23 and genetic permit approved by the Ministry of the Environment of Ecuador MAE-DNB-CM-2016-0043. We thank both institutions of the Ecuadorian State for allowing us to carry out this study. We thank CDF volunteers and researchers Mateo Reyes, Peter Pibaque, Agustín Gutiérrez, Novarino Castillo, Javier Cotín, Patricia Parker, Denis Mosquera, and Lorena Rojas, for their support in the data collection, blood sampling in the field and samples storage. Thanks Jasmine Loveland, Thibault Leroy and Ovidiu Paun from the University of Vienna, for their support in teaching data analysis and bioinformatics. Thanks to Miguel Pinto, Erwin Nemeth and Michael Dvorak for sharing their comments to improve the manuscript. This publication is contribution number 2636 of the Charles Darwin Foundation for the Galapagos Islands.

Financial support

The study was funded by a PhD fellowship awarded to David Anchundia from the World Wildlife Fund, Mr. Russell E. Train and Education for Nature, Agreement #RH94. Lakeside Fellowship awarded to David Anchundia from the California Academy of Sciences and a donation from Michael Simon to the California Academy of Sciences. Other funding sources include Galapagos Invasive Species Fund; Galapagos Conservation Trust; Galapagos Conservancy; Swiss Association of friends

of the Galapagos Islands; COmON Foundation; Re:wild; Lindblad Expeditions - National Geographic; and Peter and Kris Norvig. Grant awarded to Sabine Tebbich and Rebecca Hood-Nowotny by the Austrian Science Fund (grant number P 32555-B).

Competing interests: The authors declare no competing interests.

Data Availability

Raw sequence files used in the study are available at NCBI Sequence Read Archive, accession number SubmissionID: SUB14124997 and BioProject ID: PRJNA1064398.

Appendix A. Supplementary material data

Supplementary data to this article can be found online in the journal

References

Alström, P., Ranft, R., 2003. The use of sounds in avian systematics and the importance of bird sound archives. Bull. BOC 123, 114-135.

Anchundia, D.J., Green, R., Pike, C.L., Gutiérrez, G., Pibaque, P., Chango, R., Sevilla, C., Fessl, B., Tebbich, S., 2024a. Habitat restoration to conserve the Little Vermilion Flycatcher *Pyrocephalus nanus* on Santa Cruz Island, Galapagos. Bird Conservation International, 34, p.e14.

Anchundia, D.J., Lam, A.W., Henderson, J.B., Van Dam, M.H., Dumbacher, J.P., 2024b. Genome assembly of *Pyrocephalus nanus*, a step toward the genetic conservation of the endangered Little Vermilion Flycatcher of the Galapagos Islands. Genome Biology and Evolution, evae083.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Arbogast, B.S., Drovetski, S.V., Curry, R.L., Boag, P.T., Seutin, G., Grant, PR., Grant, R., Anderson, D.J., 2006. The origin and diversification of Galapagos mockingbirds. Evolution, 60(2), 370-382.

Baldridge, A., 1968. Stanford oceanographic expedition 17, Galapagos Islands and vicinity, 22 February-23 March 1968: observations on birds, the Galapagos fur seal, and cetaceans.

Bateson, Z.W., Dunn, P.O., Hull, S.D., Henschen, A.E., Johnson, J.A., Whittingham, L.A., 2014. Genetic restoration of a threatened population of greater prairie-chickens. Biological Conservation, 174, 12-19.

Benitez-Capistros, F., Hugé, J., Koedam, N., 2014. Environmental impacts on the Galapagos Islands: Identification of interactions, perceptions and steps ahead. Ecological Indicators, 38, 113-123.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K., Meier, R., Winker, K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution, 22(3), 148-155.

BirdLife International., 2023. *Pyrocephalus nanus*. The IUCN Red List of Threatened Species 2023: e.T103682926A172654604. https://dx.doi.org/10.2305/IUCN.UK.2023-1.RLTS.T103682926A172654604.en. Accessed on 29 March 2024.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Graham, J., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology, 15(4), e1006650.

Brekke, P., Bennett, P.M., Wang, J., Pettorelli, N., Ewen, J.G., 2010. Sensitive males: Inbreeding depression in an endangered bird. Proceedings of the Royal Society B: Biological Sciences, 277(1700) 3677-3684.

Briskie, J.V., Mackintosh, M., 2004. Hatching failure increases with severity of population bottlenecks in birds. Proceedings of the National Academy of Sciences, 101(2), 558-561.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution, 29(8), 1917-1932.

Campbell, K., Donlan, C.J., 2005. Feral goat eradication on islands. Conservation Biology, 19(5), 1362-1374.

Carmi, O., Witt, C.C., Jaramillo, A., Dumbacher, J.P., 2016. Phylogeography of the Vermilion Flycatcher species complex: Multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. Molecular Phylogenetics and Evolution, 102, 152-173.

Carpenter, A.M., Graham, B.A., Spellman, G.M., Klicka, J., Burg, T.M., 2022. Genetic, bioacoustic and morphological analyses reveal cryptic speciation in the warbling vireo complex (*Vireo gilvus*: Vireonidae: Passeriformes). Zoological Journal of the Linnean Society, 195(1), 45-64.

Causton, C.E., Peck, S.B., Sinclair, B.J., Roque-Albelo, L., Hodgson, C.J., Landry, B., 2006. Alien insects: threats and implications for conservation of Galápagos Islands. Annals of the Entomological Society of America, 99(1), 121-143.

Chaves, J.A., Parker, P.G., Smith, T.B., 2012. Origin and population history of a recent colonizer, the yellow warbler in Galápagos and Cocos Islands. Journal of Evolutionary Biology, 25(3), 509-521.

Chen, J. 2015., Github repository, https://github.com/JinfengChen/vcf-tab-to-fasta

Chen, S., Zhou, Y., Chen, Y. Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), 884-890.

Chesser, R.T., Billerman, S.M., Burns, K.J., Cicero, C., Dunn, J.L., Hernández-Baños, B.E., Jiménez, R.A., Kratter, A.W., Mason, N.A., Rasmussen, P.C., Remsen Jr, J.V., 2022. Sixty-third supplement to the American Ornithological Society's Check-list of North American Birds. Ornithology, 139(3), p.ukac020.

Colnett, J., 1798. A Voyage Into the South Atlantic and Round Cape Horn Into the Pacific Ocean. Chapter VI The Galapagoe Isles 47-61.

Cruz, F., Carrion, V., Campbell, K.J., Lavoie, C., Donlan, C.J., 2009. Bio-economics of large-scale eradication of feral goats from Santiago Island, Galapagos. The Journal of Wildlife Management, 73(2), 191-200.

Da Silva A.R., 2021. biotools: Tools for Biometry and Applied Statistics in Agricultural Science. R package version 4.2, https://cran.r-project.org/package=biotools.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., 2011. The variant call format and VCFtools. Bioinformatics, 27(15), 2156-2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience, 10(2), p.giab008.

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Molecular biology and evolution, 37(1), 291-294.

Del Hoyo, J., Collar, N.J., Christie, D.A., Elliott, A., Fishpool, L.D.C., Boesman, P., Kirwan, G.M., 2016. HBW and BirdLife International Illustrated Checklist of the Birds of the World. Volume 2: Passerines. Lynx Edicions and BirdLife International, Barcelona, Spain and Cambridge, UK.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7(1), 1-8.

Duntsch, L., Whibley, A., de Villemereuil, P., Brekke, P., Bailey, S., Ewen, J.G., Santure, A.W., 2023. Genomic signatures of inbreeding depression for a threatened Aotearoa New Zealand passerine. Molecular Ecology, 32(8), 1893-1907.

Dvorak, M., Fessl, B., Nemeth, E., Kleindorfer, S., Tebbich, S., 2012. Distribution and abundance of Darwin's finches and other land birds on Santa Cruz Island, Galapagos: Evidence for declining populations. Oryx. 46, 78-86.

Dvorak, M., Nemeth, E., Wendelin, B., Herrera, P., Mosquera, D., Anchundia, D., Sevilla, C., Tebbich, S., Fessl, B., 2017. Conservation status of landbirds on Floreana: the smallest inhabited Galapagos Island. Journal of Field Ornithology, 88(2), 132-145.

Dvorak, M., Fessl, B., Nemeth, E., Anchundia, D., Cotin, J., Schulze, C.H., Tapia, W., Wendelin, B., 2019. Survival and extinction of breeding landbirds on San Cristóbal, a highly degraded island in the Galápagos. Bird Conservation International, 30(3), 381-395.

Dvorak, M., Nemeth, E., Wendelin, B., Fessl, B., 2021. More extinctions on the Galápagos Islands? An unsuccessful search for 4 landbirds on Floreana. The Wilson Journal of Ornithology, 133(3), 514-518.

Ennos, R.A., French, G.C., Hollingsworth, P.M., 2005. Conserving taxonomic complexity. Trends in Ecology & Evolution, 20(4), 164-168.

Fessl, B., Anchundia, D., Carrion, J., Cimadom, A., Cotín, J., Cunninghame, F., Dvorak, M., Mosquera, D., Nemeth, E., Sevilla, C., Tebbich, S., Causton, C., 2017. Galapagos landbirds (passerines, cuckoos, and doves): Status, threats and knowledge gaps. In Galapagos Reports 2015-2016. GNPS, GCREG, CDF and GC. Puerto Ayora, Galapagos, Ecuador.

Fisher, A.K., Wetmore, A., 1931. Report on birds recorded by the Pinchot Expedition of 1929 to the Caribbean and Pacific. Proceedings of the United States National Museum.

FHCRC Computational Biology. Github repository. https://github.com/fhcrc/seqmagick

Frankham, R., Ballou, J.D., Ralls, K., Eldridge, M., Dudash, M.R., Fenster, C.B., Lacy, R.C., Sunnucks, P., 2017. Genetic management of fragmented animal and plant populations. Oxford University Press.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Gerdes, J.W., Gupta, S.K., Wilkerson, S.A., 2012. A review of bird-inspired flapping wing miniature air vehicle designs.

Gifford, E.W., 1919. Expedition of the California Academy of Sciences to the Galapagos Islands, 1905-1906. XIII. Field notes on the land birds of the Galapagos Islands and of Cocos Island, Costa Rica. Proc Cal Acad Sci Fourth Series, 2, 189-258.

Gould, J., 1841. Birds: Part 3 of The Zoology of the Voyage of HMS Beagle, Under the Command of Captain Fitzroy, RN, During the Years 1832 to 1836, ed. Charles Robert Darwin. London: Smith Elder and Co.

Grant, P.R., Grant, B.R., Petren, K., Keller, L.F., 2005. Extinction behind our backs: the possible fate of one of the Darwin's finch species on Isla Floreana, Galápagos. Biological Conservation, 122(3), 499-503.

Grant, B.R., Grant, P.R., 2008. Fission and fusion of Darwin's finches populations. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1505), 2821-2829.

Greenwood, P.J., Harvey, P.H., Perris, C.M., 1978. Inbreeding and dispersal in the great tit. Nature, 271, 52-54.

Hamann, O., 1981. Plant communities of the Galápagos Islands.

Hanson, C., Campbell, K., 2013. Floreana Island Ecological Restoration: Rodent and Cat Eradication Feasibility Analysis. Island Conservation: Santa Cruz, CA, USA.

Harris, M.P., 1973. The Galapagos avifauna. Condor. 75, 265-278.

Heber, S., Briskie, J.V., 2010. Population bottlenecks and increased hatching failure in endangered birds. Conservation Biology, 24(6),1674-1678.

Hedrick, P.W., Robinson, J.A., Peterson, R.O., Vucetich, J.A., 2019. Genetics and extinction and the example of Isle Royale wolves. Animal Conservation, 22(3), 302-309.

Hoeck, P.E., Bollmer, J.L., Parker, P.G., Keller, L.F., 2010. Differentiation with drift: a spatio-temporal genetic analysis of Galapagos mockingbird populations (*Mimus* spp.). Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1543), 1127-1138.

Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. Molecular ecology resources, 9(5), 1322-1332.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17(8), 754-755.

Irestedt, M., Batalha-Filho, H., Ericson, P.G., Christidis, L., Schodde, R., 2017. Phylogeny, biogeography and taxonomic consequences in a bird-of-paradise species complex, *Lophorina–Ptiloris* (Aves: Paradisaeidae). Zoological Journal of the Linnean Society, 181(2), 439-470.

Jackson, H.A., Percival-Alwyn, L., Ryan, C., Albeshr, M.F., Venturi, L., Morales, H.E., Mathers, T.C., Cocker, J., Speak, S.A., Accinelli, G.G., Barker, T., Van Oosterhout, C., 2022. Genomic erosion in a demographically recovered bird species during conservation rescue. Conservation Biology, 36(4), p.e13918.

Jamieson, I.G., Ryan, C.J., 2000. Increased egg infertility associated with translocating inbred takahe *(Porphyrio hochstetteri*) to island refuges in New Zealand. Biological Conservation, 94(1), 107-114.

Jamieson, I.G., Roy, M.S., Lettink, M., 2003. Sex-specific consequences of recent inbreeding in an ancestrally inbred population of New Zealand Takahe. Conservation Biology, 17(3), 708-716.

Johnson, H.E., Mills, L.S., Wehausen, J.D., Stephenson, T.R., Luikart, G., 2011. Translating effects of inbreeding depression on component vital rates to overall population growth in endangered bighorn sheep. Conservation Biology, 25(6), 1240-1249.

Kassambra, A., 2023. rstatix: Pipe-friendly framework for basic statistical testing (0.7.2).

Keller, L.F., 1998. Inbreeding and its fitness effects in an insular population of song sparrows (*Melospiza melodia*). Evolution, 52(1), 240-250.

Keller, L.F., Waller, D.M., 2002. Inbreeding effects in wild populations. Trends in ecology & evolution, 17(5), 230-241.

Kleindorfer, S., O'Connor, J.A., Dudaniec, R.Y., Myers, S.A., Robertson, J., Sulloway, F.J., 2014. Species collapse via hybridization in Darwin's tree finches. The American Naturalist, 183(3), 325-341.

Kruuk, L.E., Sheldon, B.C., Merilä, J., 2002. Severe inbreeding depression in collared flycatchers (*Ficedula albicollis*). Proceedings of the Royal Society of London. Series B: Biological Sciences, 269(1500), 1581-1589.

Kumar, A., 2003. Acoustic communication in birds: Differences in songs and calls, their production and biological significance. Resonance, 8, 44-55.

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.J., Wang, C., Zamani, N. Grant, B.R., 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature, 518(7539), 371-375.

Lande, R., 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. The American Naturalist, 142(6), 911-927.

Lerner, H.R., Meyer, M., James, H.F., Hofreiter, M., Fleischer, R.C., 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Current Biology, 21(21), 1838-1844.

Leuba, C., Tebbich, S., Nemeth, E., Anchundia, D., Heyer, E., Mosquera, D.A., Richner, H., Rojas Allieri, M.L., Sevilla, C., Fessl, B., 2020. Effect of an introduced parasite in natural and anthropogenic habitats on the breeding success of the endemic little vermilion flycatcher *Pyrocephalus nanus* in the Galápagos. Journal of Avian Biology, 51(8).

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Loeschcke, V., Tomiuk, J., Jain, S.K., eds., 2013. Conservation genetics (Vol. 68). Birkhäuser.

Longmire, J.L., Maltbie, M., Baker, R.J., 1997. Use of "lysis buffer" in DNA isolation and its implication for museum collections.

Matschiner, M., 2022. Ruby script update for beast 2.7 GitHub Repository https://github.com/mmatschiner/snapp_prep

Neukamm, J., Peltzer, A., Nieselt, K., 2021. DamageProfiler: fast damage pattern calculation for ancient DNA. Bioinformatics, 37(20), 3652-3653.

Nietlisbach, P., Wandeler, P., Parker, P.G., Grant, P.R., Grant, B.R., Keller, L.F., Hoeck, P.E., 2013. Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. Molecular phylogenetics and evolution, 69(3), 581-592.

Päckert, M., Martens, J., Sun, Y.H., Veith, M., 2004. The radiation of the *Seicercus burkii* complex and its congeners (Aves: Sylviidae): molecular genetics and bioacoustics. Organisms Diversity & Evolution, 4(4), 341-364.

Pérez-Pereira, N., Caballero, A., García-Dorado, A., 2022. Reviewing the consequences of genetic purging on the success of rescue programs. Conservation Genetics, 23(1), 1-17.

Picard Tools Broad Institute. GitHub Repository.

Pike, C.L., Ramirez, I.E., Anchundia, D.J., Fessl, B., Heimpel, G.E., Causton, C.E., 2021. Behavior of the avian parasite *Philornis downsi* (Diptera: Muscidae) in and near host nests in the Galapagos Islands. Journal of Insect Behavior, 34(5-6), 296-311.

Porter, D., 1822. Journal of a cruise made to the Pacific Ocean (Vol. 1). Wiley & Halsted.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller J, Sklar P, De Bakker P,I., Daly M,J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, 81(3), 559-575.

Merlen, G., 2013. Gone, gone… going: The fate of the Vermilion Flycatcher on Darwin's Islands. Galapagos Report. 2012: 180-188. GNPS, GCREG, CDF and GC, Puerto Ayora, Galapagos, Ecuador.

Mills, S., Alcántara-Rodríguez, J.A., Ciros-Pérez, J., Gómez, A., Hagiwara, A., Galindo, K.H., Jersabek, C.D., Malekzadeh-Viayeh, R., Leasi, F., Lee, J.S., Mark Welch, D.B., 2017. Fifteen species in one: deciphering the *Brachionus plicatilis* species complex (Rotifera, Monogononta) through DNA taxonomy. Hydrobiologia, 796, 39-58.

Mosquera, D., Fessl, B., Anchundia, D., Heyer, E., Leuba, C., Nemeth, E., Rojas, M.L., Sevilla, C., Tebbich, S., 2022. The invasive parasitic fly *Philornis downsi* is threatening Little Vermilion Flycatchers on the Galápagos Islands. Avian Conserv Ecol 17(1):6.

O'Connor, J.A., Sulloway, F.J., Kleindorfer, S., 2010. Avian population survey in the Floreana highlands: is Darwin's Medium Tree Finch declining in remnant patches of Scalesia forest? Bird Conservation International, 20(4), 343-353.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology, 67(5), 901-904.

Rambaut, A., 2018. FigTree version 1.4. 4. Computer program and documentation distributed by the author.

R Core Team., 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reed, D.H., Frankham, R., 2003. Correlation between fitness and genetic diversity. Conservation Biology, 17(1), 230-237.

Renteria, J.L., Gardener, M.R., Panetta, F.D., Crawley, M.J., 2012. Management of the invasive hill raspberry (*Rubus niveus*) on Santiago Island, Galapagos: eradication or indefinite control? Invasive plant science and management, 5(1), 37-46.

Rick, K., Ottewell, K., Lohr, C., Thavornkanlapachai, R., Byrne, M., Kennington, W.J., 2019. Population genomics of *Bettongia lesueur*: Admixing increases genetic diversity with no evidence of outbreeding depression. Genes, 10(11), 851.

Ridgway, R., 1894. Descriptions of twenty-two new species of birds from the Galapagos Island. Proceedings of the United States National Museum.

Ridgway, R., 1897. Birds of the Galapagos Archipelago. Proceedings of the United States National Museum.

Rothschild, L.W.R.B., Hartert, E., 1899. A review of the ornithology of the Galapagos Islands: with notes on the Webster-Harris expedition. Zoological Museum.

Rueda, D., Carrion, V., Castaño, P.A., Cunninghame, F., Fisher, P., Hagen, E., Ponder, J.B., Riekena, C.A., Sevilla, C., Shield, H., Will, D., Campbell, K.J., 2019. Preventing extinctions: planning and undertaking invasive rodent eradication from Pinzon Island, Galapagos. Island invasives: scaling up to meet the challenge.'(Eds CR Veitch, MN Clout, AR Martin, JC Russell, CJ West) 51-56.

Ruíz-Ballesteros, E., del Campo Tejedor, A., 2022. Eradicate to construct nature. Goats, rats, and humans in Floreana (Galapagos Islands). Ethnos, 1-20.

Sari, E.H., Parker, P.G., 2012. Understanding the colonization history of the Galápagos flycatcher (*Myiarchus magnirostris*). *Molecular Phylogenetics and Evolution*, *63*(2), 244-254.

Seddon, N., Tobias, J.A., 2007. Song divergence at the edge of Amazonia: an empirical test of the peripatric speciation model. Biological Journal of the Linnean Society, 90(1), 173-188.

Segelbacher, G., Bosse, M., Burger, P., Galbusera, P., Godoy, J.A., Helsen, P., Hvilsom, C., Iacolina, L., Kahric, A., Manfrin, C., Nonic, M., Buzan, E., 2022. New developments in the field of genomic technologies and their relevance to conservation management. Conservation Genetics, 23(2), 217-242.

Schofield, E.K., 1989. Effects of introduced plants and animals on island vegetation: examples from Galapagos Archipelago. Conservation Biology, 3(3), 227-239.

Schulwitz, S., Castaño, P.A., Mosquera, D., Chugcho, M., Campbell, K.J., Johnson, J.A., 2018. Floreana Island re-colonization potential of the Galápagos short-eared owl (*Asio flammeus galapagoensis*). Conservation genetics, 19, 193-205.

Snodgrass, R.E., Heller, E., 1904. Papers from the Hopkins-Stanford Galapagos Expedition, 1898-1899. XVI Birds. In Proceedings of the Washington Academy of Sciences. 231-372 Washington Academy of Sciences.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics, 22(21), 2688-2690.

Steadman, D.W., 1986. Holocene vertebrate fossils from Isla Floreana, Galápagos.

Swarth, H., 1931. The avifauna of the Galapagos Islands. Occasional papers of the California Academy of Sciences. 18, 1–299.

Toews, D.P., Irwin, D.E., 2008. Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustic analyses. Molecular Ecology, 17(11), 2691-2705.

Van Noordwijk, A.J., Scharloo, W., 1981. Inbreeding in an island population of the great tit. Evolution, 35(4), 674-688.

Venables W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.

Warrick, D.R., 1998. The turning-and linear-maneuvering performance of birds: the cost of efficiency for coursing insectivores. Canadian Journal of Zoology, 76(6), 1063-1079.

Watson, J., Trueman, M., Tufet, M., Henderson, S., Atkinson, R., 2010. Mapping terrestrial anthropogenic degradation on the inhabited islands of the Galapagos Archipelago. Oryx, 44(1), 79-82.

Willi, Y., Kristensen, T.N., Sgrò, C.M., Weeks, A.R., Ørsted, M., Hoffmann, A.A., 2022. Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. Proceedings of the National Academy of Sciences, 119(1), p.e2105076119.

Appendix A. Supplementary material

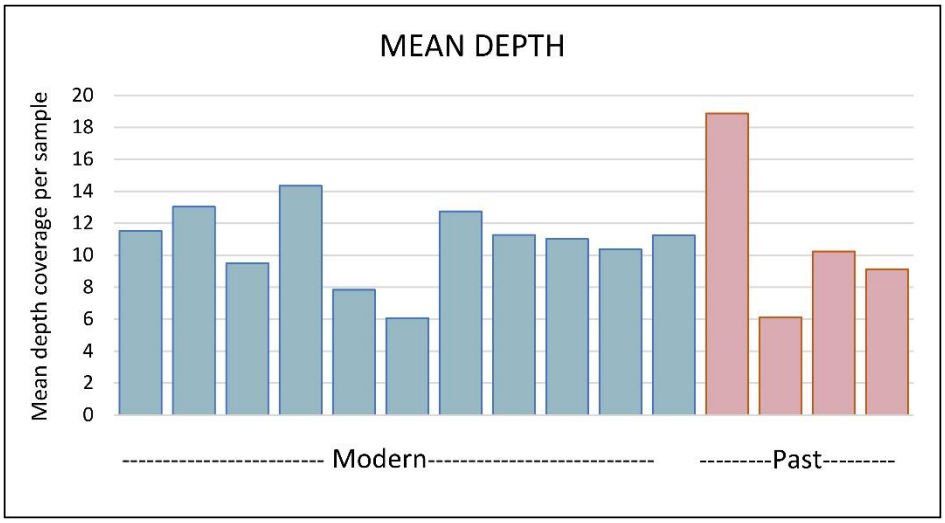


Figure S1. Average coverage of sequences of Santa Cruz individuals from the past and present.

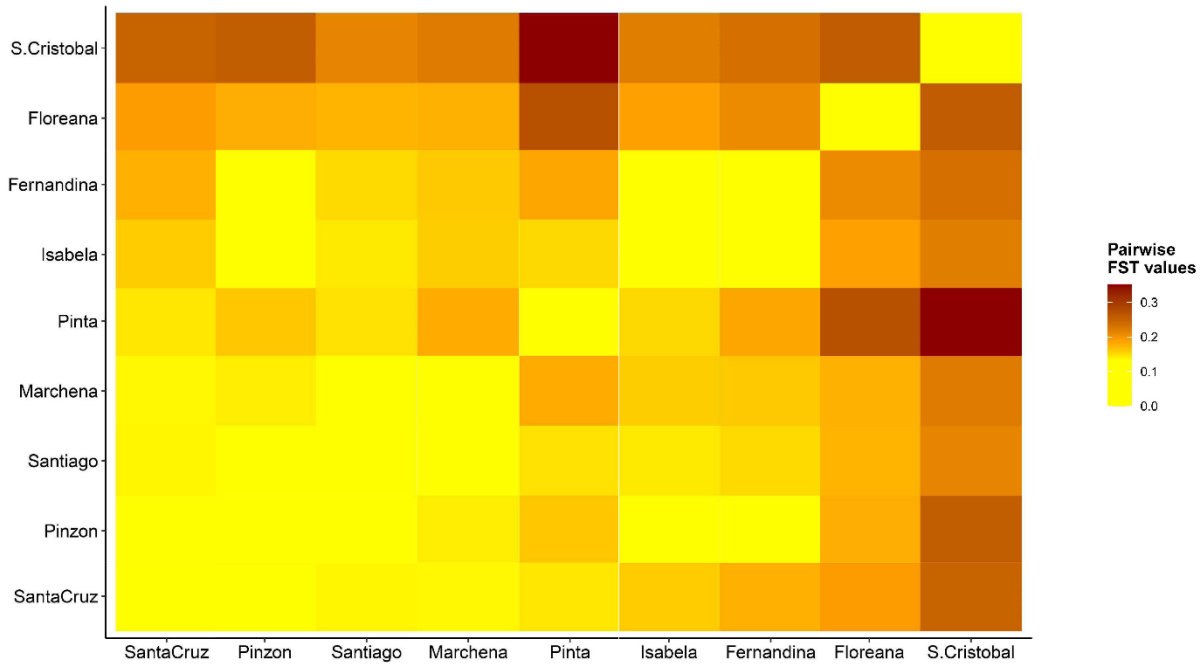


Figure S2. Graphical pairwise FST

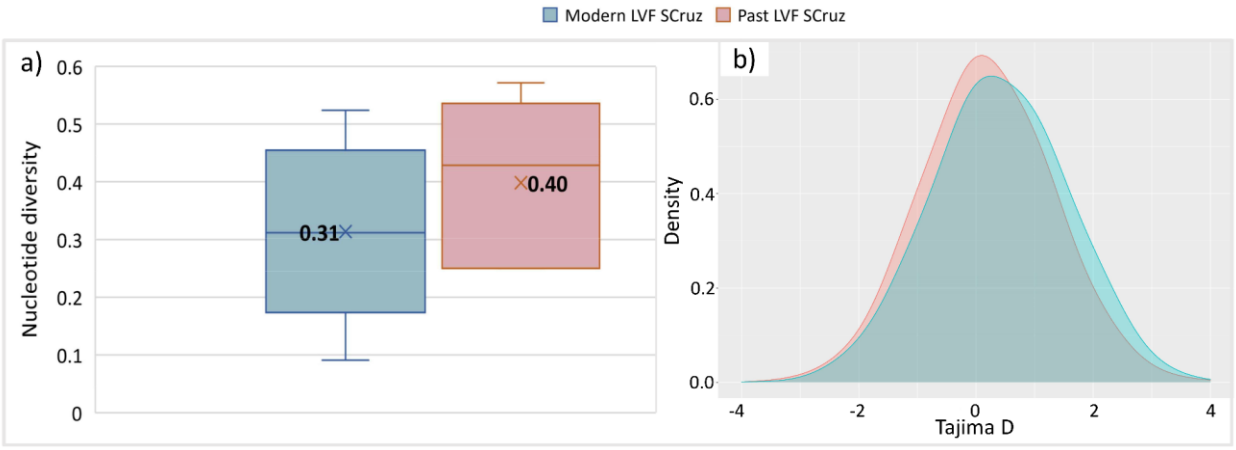
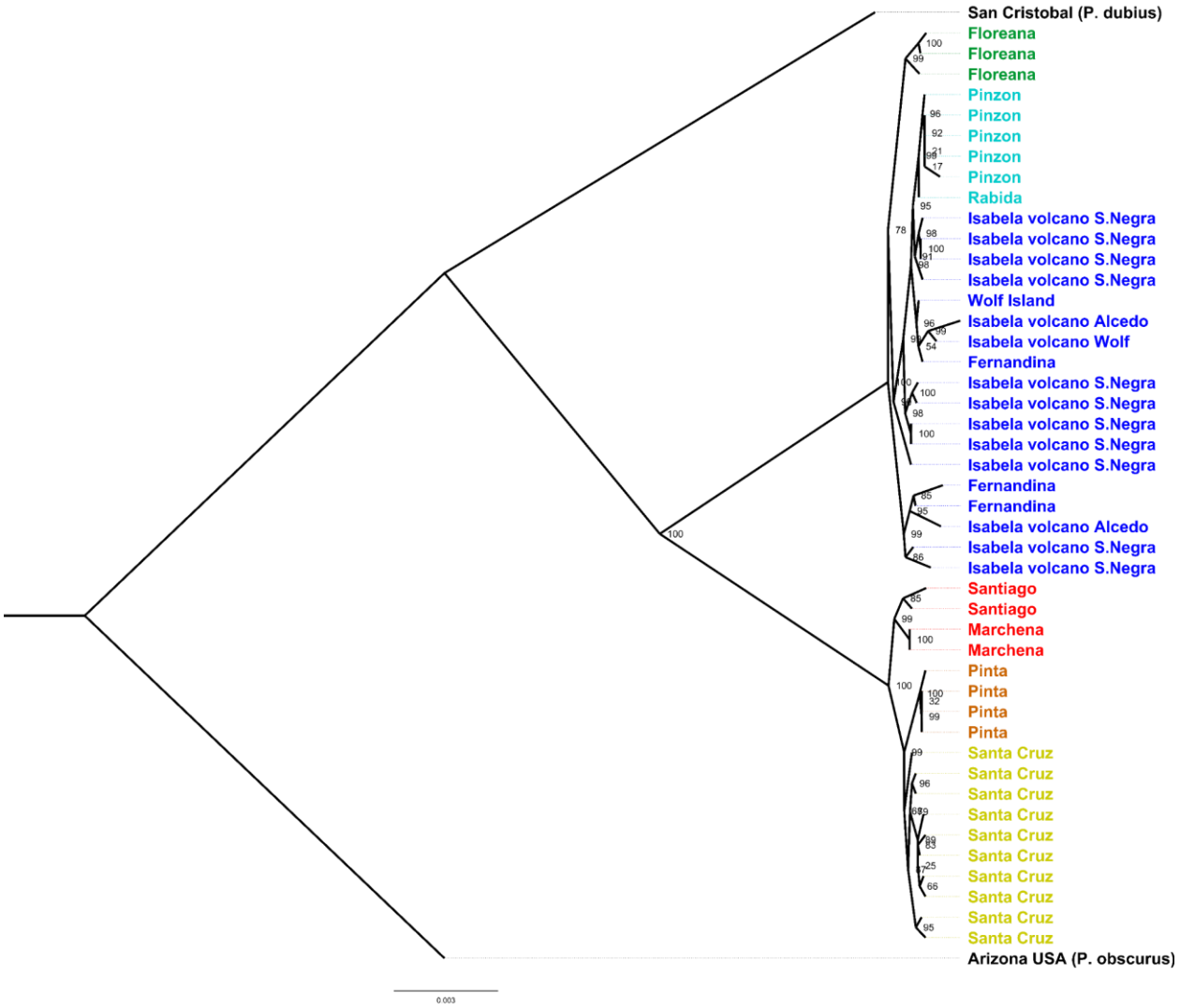


Figure S3: a) Comparison of the nucleotide diversity from birds that lived in Santa Cruz: past (1905, 1961) and present (2016 - 2021), with males and females combined (see samples list in Table 1). b) Tajima's D density plot comparison of the current population living in Santa Cruz with the past population. Results show significant differences between populations over time.



Phylogenetic ML tree built using the whole mitochondrial genome of *Pyrocephalus* with 17178 bp. The support values reported at each node are ML bootstrap values.

Table S1. Tukey HSD Post hoc test results for pairwise comparisons between islands for song variables, including A) High frequency (Hz), B) Low frequency (Hz), and C) Syllables delta time (s).

A) High Frequency

Island 1	Island 2	Mean estimate	confidence (low)	confidence (high)	p-value	significance level
Fernandina	Isabela	0.006778	-0.18852	0.202072	1	ns
	Marchena	0.481861	0.242676	0.721046	3.66E-05	****
	Pinzon	2.941825	2.70264	3.181009	2.10E-14	****
	Santa Cruz	0.332258	0.116352	0.548163	0.00117	**
Isabela	Marchena	0.475083	0.264142	0.686024	6.79E-06	****
	Pinzon	2.935047	2.724105	3.145988	2.10E-14	****
	Santa Cruz	0.32548	0.141355	0.509604	0.00022	***
Marchena	Pinzon	2.459964	2.207841	2.712087	2.10E-14	****
	Santa Cruz	-0.1496	-0.37976	0.080553	0.337	ns
Pinzon	Santa Cruz	-2.60957	-2.83972	-2.37941	2.10E-14	****

B) Low Frequency

Island 1	Island 2	Mean estimate	confidence (low)	confidence (high)	p-value	significance level
Fernandina	Isabela	0.22537	0.013581	0.43716	0.0331	*
	Marchena	-2.1259	-2.38529	-1.86651	2.11E-14	****
	Pinzon	-1.89248	-2.15187	-1.63309	2.13E-14	****
	Santa Cruz	-1.39204	-1.62618	-1.15789	5.80E-14	****
Isabela	Marchena	-2.35127	-2.58003	-2.12251	2.10E-14	****
	Pinzon	-2.11785	-2.34661	-1.88909	2.10E-14	****
	Santa Cruz	-1.61741	-1.81708	-1.41773	2.11E-14	****
Marchena	Pinzon	0.233416	-0.04	0.506835	0.12	ns
	Santa Cruz	0.733863	0.484267	0.98346	7.15E-08	****
Pinzon	Santa Cruz	0.500447	0.250851	0.750044	3.92E-05	****

C) Syllables delta time (s)

Island 1	Island 2	Mean estimate	confidence (low)	confidence (high)	p-value	significance level
Fernandina	Isabela	0.375489	-0.2047	0.955676	0.341	ns
	Marchena	1.233751	0.523169	1.944332	0.000276	***
	Pinzon	2.977094	2.266513	3.687675	6.74E-11	****
	Santa Cruz	0.607935	-0.03349	1.249356	0.0692	ns
Isabela	Marchena	0.858261	0.231588	1.484935	0.00401	**
	Pinzon	2.601604	1.974931	3.228278	8.15E-11	****
	Santa Cruz	0.232446	-0.31456	0.779451	0.722	ns
Marchena	Pinzon	1.743343	0.994325	2.492361	4.02E-06	****

	Santa Cruz	-0.62582	-1.30957	0.057941	0.0842	ns
Pinzon	Santa Cruz	-2.36916	-3.05292	-1.6854	3.15E-09	****

General discussion

Pyrocephalus nanus - reproduction in Santa Cruz

Habitat degradation on Santa Cruz Island significantly impacts the reproductive success of *P. nanus* (Anchundia et al., 2004a). My results confirm that habitat deterioration and the consequent reduction of access to food sources are primary factors causing females to abandon their nests during the early stages of incubation. This adverse effect occurs before the Avian Vampire Fly parasitizes the nests. Due to this, *P. nanus* faces two major sequential challenges, such that an initial limited food availability yields a lower energy capacity of the birds for nest maintenance, resulting in nest abandonment. However, when food was accessible, the birds continued to incubate their nests; however, the Avian Vampire Fly subsequently parasitized the nests within a few days. Our findings indicate that addressing both issues—enhancing food availability and controlling parasitism—improves the reproductive success of *P. nanus*, allowing them to spend more time for nest brooding and offspring care.

Additionally, we observed and confirmed that *P. nanus* thrived better in open places. This was somewhat expected, since flycatchers generally prefer these types of open areas (Beedy, 1981; Mannan, 1984). Unfortunately, invasive vegetation on Santa Cruz has significantly altered the ecosystem (Rentería et al., 2012; Walentowitz et al., 2021) and the absence of herbivores has allowed invasive plant species to expand. A pressing concern is that the *P. nanus* population in Santa Cruz is very small and the genetically effective population size is even smaller because there are more males than females. The survival of this population relies heavily on the reproductive success of the relatively few females of reproductive age (~11 individuals). Thus, it is crucial to continue and expand restoration efforts to enhance breeding conditions for more pairs, thereby preventing the extinction/extirpation of *P. nanus* on Santa Cruz.

Another critical aspect of habitat restoration is the management of nest parasitism by the Avian Vampire Fly. The insecticide injection method, which is a short-term solution to control parasitism, successfully eliminated these nest parasites in my study, similar to previous studies with landbirds in Galapagos (Cimadom et al., 2019; Leuba et al., 2020). However, a long-term solution currently in development using biological control to reduce the prevalence of these parasites (Boulton et al., 2019; Ramirez, 2023). By addressing both the food scarcity and parasitism issues, habitat restoration creates a more favourable environment for vermilion flycatchers in Santa Cruz and potentially on other islands where the populations have declined.

Moreover, habitat restoration may be facilitated by the reintroduction of native herbivores that might play a vital role in controlling invasive plant species. The largest native herbivores in Santa Cruz Island are two species of giant tortoises (*Chelonoidis porteri, Chelonoidis donfaustoi*)

(Poulakakis et al., 2015). Giant tortoises in Galapagos were decimated by whalers, buccaneers, early expeditions, and impacted by invasive predators, etc, (Garrick et al., 2015), with an estimated 100,000 to 200,000 tortoises harvested on several islands over two centuries (MacFarland et al., 1974; Conrad and Gibbs, 2021). In Santa Cruz, both species currently show signs of inbreeding due to drastic population declines (Garrick et al., 2015). One of these species, *C. donfaustoi*, has an estimated population size of only 250 individuals (Poulakakis et al., 2015) while *C. porteri* has an estimated population size of ~3400 individuals (Cayot et al., 2007).

Currently, tortoises are not reported to be present in the center of Santa Cruz (Poulakakis et al., 2015, Bastille-Rousseau et al., 2017) where the last remnant of *P. nanus* remains on this island. However, it is likely that tortoises were more widespread on Santa Cruz Island in the past. These large herbivores have an important role in transforming ecosystems in the Galapagos (Hunter et al., 2021). Therefore, their reintroduction to the highlands may help re-establish the ecosystem balance and prevent the overgrowth of invasive plants; however, more studies are necessary before moving giant tortoises to the location where *P. nanus* is still present. Importantly, it is necessary to determine which of the two species was formerly present in that area before potentially conducting translocations.

Genetic diversity of P. nanus population in Santa Cruz

To obtain genetic data, I randomly sampled one-third of the *P. nanus* population from Santa Cruz from 2019-2020. Analysis of this data revealed no signs of inbreeding depression or excess homozygosity, despite the population's small size. Contrary to our expectations, the analysis indicated that paired individuals were not closely related, suggesting there was limited inbreeding within this small population. Data from before habitat restoration (2017-2019) indicated that breeding attempts were largely unsuccessful, leading to low recruitment. Additionally, I postulate that prior to our monitoring in 2017, breeding rates were similarly low because the population decreased very quickly in just a few years. It is also important to note that the genetic samples used in my study are from birds that lived from 2019-2020, a period marked by minimal reproductive success. Thus, it is unclear whether these findings reflect the current state of the Santa Cruz population in 2024.

Since 2021, habitat restoration efforts have resulted in higher reproductive success each breeding season, albeit with a limited number of breeding pairs (Anchundia et al., 2004a). Most individuals sampled in 2019-2020 have most likely died as the lifespan of these birds is short (5-6 years long), with only their descendants remaining to maintain the population. If habitat restoration had not taken place, the Santa Cruz population would likely have faced extinction in coming years. The surviving individuals are descendants of individuals that bred during the initial habitat restoration efforts during my study. Their offspring are now breeding amongst themselves, raising new concerns

about potential inbreeding within the population in 2024. Ongoing genetic assessments in the coming years are essential for monitoring and managing their population dynamics.

Analysis of ancient DNA from *P. nanus* on Santa Cruz from 60 and 115 years ago provided a window into the past genetics of this population. While we did not have a large sample size from the past to compare with birds that lived in modern times in Santa Cruz, the large number of SNPs (N=233,916) obtained from these five skin museums were sufficient to draw several conclusions. Ancient DNA analysis revealed a higher nucleotide diversity in the past compared to the current population in Santa Cruz, with a significant decline in diversity of 22.5%. Surprisingly, we found in the museum collection that *P. nanus* lived on Baltra Island and in the arid zone of Santa Cruz, areas currently completely devoid of these birds. (California Academy of Sciences ORN:3211, 3087). This suggests *P. nanus* from Santa Cruz has the capabilities to thrive in drier conditions too, but that some yet unidentified factor affected their survival in these areas. The decline in genetic diversity could have implications for adaptability of these birds. However, the history of Galapagos species colonization suggests that small founding populations can overcome challenges through adaptation and expansion, (Grant et al., 2001, Hedrick, 2019) highlighting the resilience of the ancestors of this group. Therefore, it is possible that *P. nanus* can recover in Santa Cruz if they have more favorable habitat conditions and protection from parasitism by the Avian Vampire Fly.

Genome assembly

The assembly of the *P. nanus* genome enables more comprehensive studies on population genetics and phylogeography in the Galapagos to be conducted. This genome holds particular significance within the Tyrannidae bird family, where assembled genomes are scarce, making it a valuable resource for broader research on lesser-studied species (Anchundia et al., 2024b). Moreover, it underscores the importance of conducting further investigations on this genus, not only in the Galapagos but also across the American continent. The complete genome is essential for identifying SNPs that are under selection or that are involved in local adaptation and identify the genes or mechanisms involved (Brandies et al., 2019, Fuentes-Pardo and Ruzzante, 2017). In addition, it is important to obtain transcriptomic data to aid with the detailed annotation. This genome offers opportunities for studies on how birds adapt to diverse habitats and climates (Wang et al., 2022). Such insights are crucial for understanding their capacity to thrive and potentially inform conservation strategies, including translocations and habitat management.

Phylogenetics of Pyrocephalus in Galapagos

Phylogenetic analyses shed light on the colonization history and current relationships among *Pyrocephalus* populations in the Galapagos archipelago. Surprisingly, this research has identified the Santa Cruz population as a distinct evolutionary unit with a unique genetic lineage, highlighting the urgent need for its conservation. Contrary to prior assumptions, *P. nanus* and *P. dubius* have been revealed as old groups in the Galapagos, predating other well-known endemic species such as mockingbirds, Galapagos flycatchers, ground finches, and tree finches (Nietlisbach et al., 2013; Arbogast et al., 2006; Lamichhaney et al., 2015; Sari and Parker, 2012). My study has identified at least seven distinct genetic groups within the *Pyrocephalus* genus, supported by genetic, acoustic, and morphological characters. Despite comprehensive analysis of the data that I had, I also found evidence of extinct populations from Santa Fe Island (four museums specimens collected in 1900 and housed at the American Museum of Natural History), which suggests future work to understand the phylogenetic affinities of birds from Santa Fe. Furthermore, the sampling of birds from Marchena and Santiago included only two birds per island, so additional work for these islands may elucidate genetic interactions and patterns of migration and colonization.

Taxonomy

Current taxonomic classifications by Birdlife International and Clements Checklist of Birds of the World recognize the presence of endemic *Pyrocephalus* species in the Galapagos, although discrepancies exist in the number of species and subspecies identified (Clements et al., 2023; Birdlife International, 2024). Birdlife accepts two species *P. nanus* and *P. dubius* (Birdlife International, 2024), while Clements recognizes one species with two subspecies (*P. nanus nanus* and *P. nanus dubius* (Clements et al., 2023). However, the distinct differences within *P. nanus* observed in this study warrant a reevaluation of its taxonomic status, potentially leading to a refined classification that aligns with the unique evolutionary trajectories of distinct island populations. Certain island populations of *Pyrocephalus* no longer mix and are on their own evolutionary trajectory. Studies of isolation on islands show that when there is no admixture with other populations, populations take their own evolutionary course (Giarla et al., 2018; Losos and Ricklefs, 2009), which eventually results in the formation of new species. Therefore, isolation is one of the main requirements for a new species to appear (MacArthur and Wilson, 1963; Hamilton and Rubinoff, 1967). These data support arguments for *Pyrocephalus* reclassification in Galapagos and highlight its conservation. Emphasis should be placed on research efforts dedicated to each distinct group, because these genetic lineages are unique and these island populations are declining very quickly (Fessl et al., 2017).

Translocation

One distinct population of *P. nanus* is from Isabela, which is currently the largest living *Pyrocephalus* population in Galapagos (CDF, unpublished data). Its isolation for over 0.7 MYA, despite its close proximity to several islands, and no evidence of movement toward this island recently is quite surprising. Island biography theories predict that the number of species and population size on an island increases with increasing island size (MacArthur and Wilson, 1967), and it is less likely for a species to become extinct on a large island than on a small island (MacArthur and Wilson, 1967; Losos and Ricklefs, 2009). The larger size of Isabela Island most likely contributes to the resilience of *P. nanus* there and hopefully continues to support a larger population.

Currently, a management program has been developed that plan to reintroduce *Pyrocephalus* birds to Floreana Island. We suggest that this reintroduction should be conducted after addressing and eliminating the problems that contributed to the past extirpation of *Pyrocephalus* on Floreana. As this population is already extirpated, there would be no concerns of genetic matching or lineage mixing; therefore, managers could select birds that would best adapt to Floreana's ecological conditions. Perhaps the most ideal populations are those on Isabela, which has the most abundant populations in the archipelago and has both arid and humid ecosystems like Floreana. Similar to that planned for Floreana Island, there is another management program in place to reintroduce *Pyrocephalus* to Santa Fe Island. The last sighting of *Pyrocephalus* birds on this island was in 1929 (Fisher and Wetmore, 1931). Therefore, we still do not know the genetics of this population, and thus, whether another distinct lineage was lost. However, since this population is no longer present on the island, a different lineage with an ecologically similar habitat could be used for reintroduction, such as that from Pinzon Island. Overall, a more detailed study is necessary to evaluate a reintroduction of *Pyrocephalus* to Santa Fe.

Conclusion

Finally, in modern times, remote island archipelagos are no longer isolated due to the extensive movement of humans. This movement provides more opportunities for arrivals of new species, which can become invasive and negatively affect their novel environments, such as by degrading the habitat. Given my findings from my PhD thesis and the current status of *Pyrocephalus* populations in Galapagos, I recommend the following actions: 1) Habitat restoration: Continue and expand habitat restoration efforts on Santa Cruz Island to enhance breeding conditions for *Pyrocephalus nanus* and focus on removing invasive vegetation, creating open spaces, and improving food availability, 2) Control nest parasitism from the Avian Vampire Fly: Implement short-term solutions until long-term solutions are available, 3) Reintroduce native herbivores: Conduct further studies to evaluate the

feasibility and impact of reintroducing native herbivores, particularly giant tortoises, to the habitat of *P. nanus* on Santa Cruz and assess their role in controlling invasive plant species and restoring ecosystem balance, 4) Ongoing monitoring: Continuously monitor the genetic dynamics and reproductive success of the *P. nanus* population on Santa Cruz Island and assess potential inbreeding and manage genetic risks through proactive interventions, 5) Taxonomic revision: Based on the genetic analysis and distinct evolutionary units identified, reevaluate the taxonomic classification of *Pyrocephalus* species in the Galapagos and consider a refined classification that aligns with the unique evolutionary trajectories of distinct island populations, 6) Further research: Conduct additional research to fill knowledge gaps, such as obtaining specimens from Santa Fe Island and investigating functions of the genes for adaptations to difference habitats, and 7) Education: Raise public awareness about the conservation needs of *Pyrocephalus* species in the Galapagos and educate locals and visitors about the importance of protecting these unique birds and their habitats.

REFERENCES

Anchundia, D.J., Green, R., Pike, C.L., Gutiérrez, G., Pibaque, P., Chango, R., Sevilla, C., Fessl, B., Tebbich, S., 2024a. Habitat restoration to conserve the Little Vermilion Flycatcher *Pyrocephalus nanus* on Santa Cruz Island, Galapagos. *Bird Conservation International*, 34, p.e14.

Anchundia, D.J., Lam, A.W., Henderson, J.B., Van Dam, M.H., Dumbacher, J.P., 2024b. Genome assembly of *Pyrocephalus nanus*, a step toward the genetic conservation of the endangered Little Vermilion Flycatcher of the Galapagos Islands. *Genome Biology and Evolution*, evae083.

Arbogast, B.S., Drovetski, S.V., Curry, R.L., Boag, P.T., Seutin, G., Grant, PR., Grant, R. and Anderson, D.J. (2006). The origin and diversification of Galapagos mockingbirds. *Evolution*, 60(2), 370-382.

Brandies, P., Peel, E., Hogg, C.J. and Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, *10*(11), 846.

Bastille-Rousseau, G., Gibbs, J.P., Yackulic, C.B., Frair, J.L., Cabrera, F., Rousseau, L.P., Wikelski, M., Kümmeth, F. and Blake, S. (2017). Animal movement in the absence of predation: environmental drivers of movement strategies in a partial migration system. *Oikos*, *126*(7), 1004-1019.

Beedy, E.C. (1981). Bird communities and forest structure in the Sierra Nevada of California. *Condor*, *83*(2), 97-105.

BirdLife International. (2024). Species factsheet: Little Vermilion flycatcher *Pyrocephalus nanus* Downloaded from https://datazone.birdlife.org/species/factsheet/little-vermilion-flycatcher-pyrocephalus-nanus on 30/08/2024.

Boulton, R.A., Bulgarella, M., Ramirez, I.E., Causton, C.E. and Heimpel, G.E. (2019). Management of an invasive avian parasitic fly in the Galapagos Islands: is biological control a viable option. *Island invasives: scaling up to meet the challenge. IUCN, Gland*. 360-363.

Cayot, L.J., Gibbs, J.P., Tapia, W. and Caccone, A. (2017). *Chelonoidis porteri*. *The IUCN Red List of Threatened Species* 2017: e.T9026A82777132. https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T9026A82777132.en. Accessed on 14 August 2024.

California Academy of Sciences. CAS Ornithology (ORN). Record ID: urn:catalog:CAS:ORN:3211. Source: http://ipt.calacademy.org:8080/ipt/resource.do?r=orn (source published on 2019-07-15)

California Academy of Sciences. CAS Ornithology (ORN). Record ID: urn:catalog:CAS:ORN:3087. Source: http://ipt.calacademy.org:8080/ipt/resource.do?r=orn (source published on 2019-07-15)

Cimadom, A., Jäger, H., Schulze, C. H., Hood-Nowotny, R., Wappl, C. and Tebbich, S. (2019). Weed management increases the detrimental effect of an invasive parasite on arboreal Darwin's finches. *Biological Conservation,* 233, 93-101.

Clements, J.F., Rasmussen, P.C., Schulenberg, T.S., Iliff, M.J., Fredericks, T.A., Gerbracht, J.A., Lepage, D., Spencer, A., Billerman, S.M., Sullivan, B.L. and Wood, C.L. (2023). The eBird/Clements Checklist of Birds of the World: V2023. Cornell University, New York, USA.

Conrad, C. and Gibbs, J.P. (2021). The era of exploitation: 1535–1959. *In Galapagos giant tortoises* (pp. 63-81). Academic Press.

Fessl, B., Anchundia, D., Carrion, J., Cimadom, A., Cotín, J., Cunninghame, F., Dvorak, M., Mosquera, D., Nemeth, E., Sevilla, C., Tebbich, S., and Causton, C. (2017). Galapagos landbirds (passerines, cuckoos, and doves): Status, threats and knowledge gaps. In Galapagos Reports 2015-2016. GNPS, GCREG, CDF and GC. Puerto Ayora, Galapagos, Ecuador.

Fisher, A.K. and Wetmore, A. (1931). Report on birds recorded by the Pinchot Expedition of 1929 to the Caribbean and Pacific. *Proceedings of the United States National Museum*.

Fuentes-Pardo, A.P and Ruzzante, D.E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular ecology*, *26*(20), 5369-5406.

Garrick, R.C., Kajdacsi, B., Russello, M.A., Benavides, E., Hyseni, C., Gibbs, J.P., Tapia, W. and Caccone, A. (2015). Naturally rare versus newly rare: demographic inferences on two timescales inform conservation of Galápagos giant tortoises. *Ecology and Evolution*, *5*(3), 676-694.

Giarla, T.C., Maher, S.P., Achmadi, A.S., Moore, M.K., Swanson, M.T., Rowe, K.C., Esselstyn, J.A. (2018). Isolation by marine barriers and climate explain areas of endemism in an island rodent. *Journal of Biogeography*, *45*(9), 2053-2066.

Grant, P.R., Grant, B.R., and Petren, K. (2001). A population founded by a single pair of individuals: establishment, expansion, and evolution. *Microevolution rate, pattern, process*, 359-382.

Hamilton, T.H. and Rubinoff, I. (1967). On predicting insular variation in endemism and sympatry for the Darwin finches in the Galapagos archipelago. *The American Naturalist*, *101*(918), 161-171.

Hedrick, P.W. (2019). Galapagos Islands endemic vertebrates: a population genetics perspective. *Journal of Heredity*, *110*(2), 137-157.

Hunter, E.A., Blake, S., Cayot, L.J. and Gibbs, J.P. (2021). Role in ecosystems. In *Galapagos giant tortoises* (pp. 299-315). Academic Press.

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.J., Wang, C., Zamani, N. Grant, B.R. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371-375.

Leuba, C., Tebbich, S., Nemeth, E., Anchundia, D., Heyer, E., Mosquera, D.A., Richner, H., Rojas Allieri, M.L., Sevilla, C. and Fessl, B. (2020). Effect of an introduced parasite in natural and anthropogenic habitats on the breeding success of the endemic little vermilion flycatcher *Pyrocephalus nanus* in the Galápagos. *Journal Avian Biology, 51*(8), e02438.

Losos, J.B. and Ricklefs, R.E. (2009). Adaptation and diversification on islands. *Nature*, *457*(7231), 830-836.

MacArthur, R.H. and Wilson, E.O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, 373-387.

MacArthur, R.H. and Wilson, E.O. (1967). The Theory of Island Biogeography. *Princeton University Press*.

MacFarland, C.G., Villa, J. and Toro, B. (1974). The Galápagos giant tortoises (*Geochelone elephantopus*) Part I: Status of the surviving populations. *Biological Conservation, 6*(2),118-133.

Mannan, R.W. and Meslow, E.C. (1984). Bird populations and vegetation characteristics in managed and old-growth forests, northeastern Oregon. *The Journal of Wildlife Management, 48*(4), 1219-1238.

Nietlisbach, P., Wandeler, P., Parker, P.G., Grant, P.R., Grant, B.R., Keller, L.F. and Hoeck, P.E. (2013). Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. *Molecular Phylogenetics and Evolution*, *69*(3), 581-592.

Poulakakis, N., Edwards, D.L., Chiari, Y., Garrick, R.C., Russello, M.A., Benavides, E., Watkins-Colwell, G.J., Glaberman, S., Tapia, W., Gibbs, J.P. and Cayot, L.J. (2015). Description of a new Galápagos giant tortoise species (Chelonoidis; Testudines: Testudinidae) from Cerro Fatal on Santa Cruz island. *PLoS One*, *10*(10), e0138779.

Ramirez, I. (2023). Wasps, vampires, and carrion flies: addressing the safety of the parasitoid *Conura annulifera* (Hymenoptera: Chalcididae), a promising biological control agent for the Galapagos Islands.

Rentería, J.L., Gardener, M.R., Panetta, F.D., Atkinson, R. and Crawley, M.J. (2012). Possible impacts of the invasive plant *Rubus niveus* on the native vegetation of the Scalesia forest in the Galapagos Islands. *PLoS One, 7*(10), e48106.

Sari, E.H. and Parker, P.G. (2012). Understanding the colonization history of the Galápagos flycatcher (*Myiarchus magnirostris*). *Molecular Phylogenetics and Evolution*, *63*(2), 244-254.

Walentowitz, A., Manthey, M., Bentet Preciado, M.B., Chango, R., Sevilla, C. and Jäger, H. (2021). Limited natural regeneration of unique *Scalesia* forest following invasive plant removal in Galapagos. *PloS One*, *16*(10), e0258467.

Wang, J., Yu, J., Lipka, A.E., and Zhang, Z. (2022). Interpretation of Manhattan plots and other outputs of genome-wide association studies. In *Genome-Wide Association Studies* (pp. 63-80). New York, NY: Springer US.

Summary

The Little Vermilion Flycatcher (LVF), *Pyrocephalus nanus*, from Galapagos has suffered drastic population declines on Santa Cruz Island, with only 30-40 individuals remaining. We assessed the impact of habitat management on its breeding success and examined its genetic diversity using whole genomes to understand gene flow between islands and identify the most closely related population for potential translocations. We experimentally removed invasive plants species from seven plots and compared them with unmanaged areas, measuring incubation time, daily failure rate (DFR), breeding success, foraging, and perch height. In 2021, reproductive success was significantly higher in managed areas, with a notable reduction in DFR during the egg and nestling stages. Females incubated significantly more in managed areas, while foraging and perch heights were lower compared to unmanaged areas. To assess genetic diversity, we sequenced genomes from birds collected in 1905, 1961, and 2020-21. We found a significant decline in nucleotide diversity, but no evidence of inbreeding. Phylogenetic analysis using 233,916 SNPs revealed that *Pyrocephalus* in the Galápagos is much older than previously thought, with an estimated arrival 1.2 million years ago. Birds from Pinzón Island were identified as the closest population for potential translocation to Santa Cruz. Finally, we found that *P. nanus* diverges into five distinct genetic lineages, which could eventually be classified as separate subspecies or species.

Zusammenfassung

Der Zwergschnäpper, *Pyrocephalus nanus*, von den Galapagosinseln hat auf der Insel Santa Cruz drastische Populationsrückgänge erlitten, wobei nur noch 30-40 Exemplare übrig sind. Wir haben die Auswirkungen der Lebensraumbewirtschaftung auf seinen Bruterfolg bewertet und seine genetische Vielfalt anhand ganzer Genome untersucht, um den Genfluss zwischen den Inseln zu verstehen und die am engsten verwandte Population für mögliche Umsiedlungen zu identifizieren. Wir haben experimentell invasive Pflanzenarten aus sieben Parzellen entfernt und sie mit nicht bewirtschafteten Gebieten verglichen, wobei wir Inkubationszeit, tägliche Ausfallrate (TA), Bruterfolg, Nahrungssuche und Ansitzhöhe gemessen haben. Im Jahr 2021 war der Fortpflanzungserfolg in bewirtschafteten Gebieten deutlich höher, mit einer deutlichen Verringerung der TA während der Ei- und Nestlingsphase. Die Weibchen brüteten in bewirtschafteten Gebieten deutlich mehr, während Nahrungssuche und Ansitzhöhe im Vergleich zu nicht bewirtschafteten Gebieten niedriger waren. Um die genetische Vielfalt zu bewerten, haben wir Genome von Vögeln sequenziert, die 1905, 1961 und 2020-21 gesammelt wurden. Wir stellten einen deutlichen Rückgang der Nukleotidvielfalt fest, aber keine Hinweise auf Inzucht. Eine phylogenetische Analyse anhand von 233.916 SNPs ergab, dass *Pyrocephalus* auf den Galapagosinseln viel älter ist als bisher angenommen. Seine Ankunft wurde auf 1,2 Millionen Jahre geschätzt. Vögel von der Insel Pinzón wurden als die nächstgelegene Population für eine mögliche Umsiedlung nach Santa Cruz identifiziert. Schließlich stellten wir fest, dass *P. nanus* in fünf verschiedene genetische Linien divergiert, die schließlich als separate Unterarten oder Arten klassifiziert werden könnten.