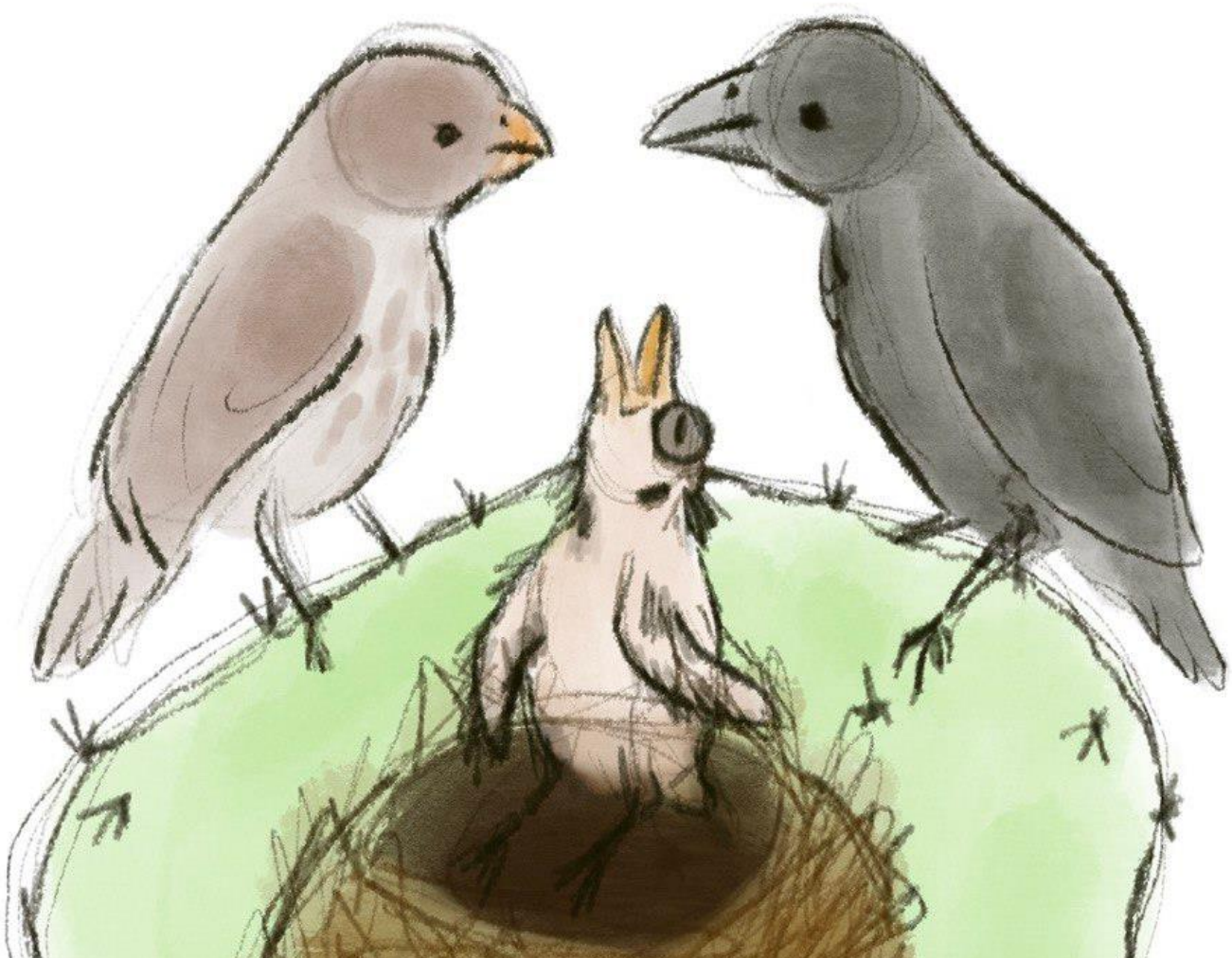
Lorenzo Micolucci

A novel microbiome-based approach to assess hybridization of cactus finches (*Geospiza scandens*) in Jardín de Opuntia, Galápagos

30th of November 2023

Supervised by: Jente Ottenburghs

Wildlife Ecology and Conservation
WEC80436



Acknowledgements

Working on this thesis has been very challenging, from the ever-changing research planning, the harsh decisions due to material and technical limits, to the final writing marathon. There are many people I need to thank for helping me deal with all of this problems and supporting me. First of all, I would like to thank my supervisor Jente Ottenburghs who has always been very kind and understandable in my occasional doubts and struggles. I'm incredibly thankful to Jaime Chaves not only for the opportunity he gave me to join his research team on the wonderful Galápagos islands, but also for shaping me into a novice ornithologist. I'd like to thank Alice Dalla Pietà for being a patient research companion and a friend, in the good times and in the bad times. Thank you Paul, Kerly, Naza, Diego, Diana and José for all of the selfless help at the GSC laboratories. Thank you Kiyoko Gotanda, Dominique Potvin and Sarah Knutie for your expertise, tips and help which made sure the research could always continue. Last but not least, I'm beyond thankful to my statistics and programming guru, Pippa.

Table of contents

0 Summary

Research on microbiome has become very popular in recent years due to discoveries about its influence on many biological factors. In this thesis, a novel way of identifying hybrids is proposed based on gut microbial data. Since cactus finches (*Geospiza scandens*) and small ground finches (*Geospiza fuliginosa*) feed their young through regurgitation, the microbiome of a hybrid is hypothesized to resemble both parental species. To test for this hypothesis, beak morphology measurements and feces were analyzed (Figure 0). From fecal samples bacterial DNA was extracted and sequenced using Nanopore's MinION. Results show how morphological data suggests the existence of three distinct morphotypes in the population (two parental species and the possible hybrids). Microbial data at genus level instead does not correspond to any difference between cactus, small ground and supposed hybrid finches. The hypothesis is confirmed since all three groups appear similar, but the novel microbiome-based approach remains untested. Gut microbial compositions may then be defined by the birds' diet which could be overlapping resulting in indistinguishable microbiomes. Furthermore, bacterial genus identification may not be informative enough and future studies are suggested to implement more detailed microbiome analyses at a species level.

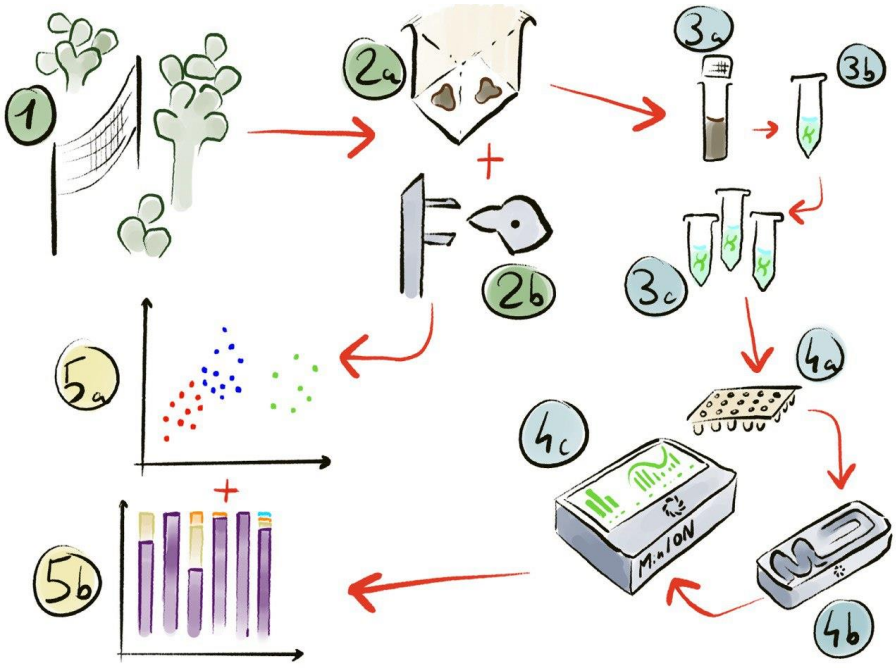


Figure 0: Visual experimental summary. Fieldwork steps in green, laboratory work in blue and data analyses in yellow. 1: Finches are captured through mist netting in Jardin de Opuntia. 2a: Finches feces are collected using a feces collection apparatus; 2b: Beak morphology is measured. 3: from feces (3a) extraction on bacterial DNA (3b) and amplification through PCR (3c). 4: Library preparation and barcode association (4a), barcode pooling in flow-cell (4b), sequencing with MinION (4c). 5a: representation of results from morphology analyses. 5b: representation of microbiome based results.

1 Introduction

1.1 Microbiome

Microbiome studies have had a surge in popularity in the recent years (Cullen *et al*. 2020). This increased interest in the topic derives from a series of discoveries on the intestinal flora's intrinsic connection with its host. The influence of the microbiota not only can be seen on its hosts physiology but also its development (Sommer *et al*. 2013), behaviour (Morais *et al*. 2021) and, on a larger scale, even in its evolutionary history (Henry *et al*. 2021).

Most vertebrates harbor a gut microbiome with varying biodiversity compositions influenced by the animal's dietary habits (Hammer *et al.* 2019). The microorganisms are mostly bacteria but can also be fungi, archaea, protists or algae (Ley *et al*. 2008). Microbial populations don't share DNA with their host and as such are not always vertically inherited. The intestines of a newly born organisms can be colonized by their immediate post-natal environment and by passing through the birth route (Miller *et al* 2021). Throughout development, later stages of microbiome acquisition involve transmission events including dietary factors and the organism's habitat (Miller *et al* 2021). Microbes present in the post-natal and developmental environment are often belonging to the newborns' parents and as such, a degree of vertical inheritance is often present (Miller *et al* 2021).

1.2 Hybridization

Microbiome has also been investigated recently as a possible marker in hybridizing animals (Miller *et al* 2021). Hybridization is a phenomenon that has been described across the whole Eukaryote domain and it can be defined as mating between morphologically and genetically distinguishable biological species (Short 1969, Arnold 1997).

Multiple studies on mammals have demonstrated how microbiome varies in hybrid organisms (Miller *et al* 2021). In a study on elk hybrids of red deer (*Cervus elaphus*) and sika deer (*Cervus nippon*), the abundance of specific symbiotic rumen bacteria has been shown to be higher than in the two parents (Li *et al*. 2016). Research investigating mating between two different mice subspecies (*Mus musculus musculus* and *Mus musculus domesticus*) found that hybrids possess a greater gut microbial diversity compared to their parental subspecies taken individually (Wang *et al*. 2015).

In avian species, hybridization has been demonstrated to be difficult to determine (Randler 2004, Ottenburghs 2023). Generally, two main different approaches in hybrid detection are taken utilizing both morphology and genetics. First a morphological assessment is made but that has been shown to be prone to error (Randler 2004, Ottenburghs 2023). Hybrids often appear as an intermediate morphotype in relation to the original two species. Given the inherent limitations of morphological identification, genetic analyses serve as a valuable adjunct for more precise hybrid detection. Genetic investigations can be done at many levels: on whole genomes, individual loci (Lamichhaney *et al*. 2018, Kong *et al*. 2021), using mitochondria or microsatellite markers (Helbig *et al*. 2001).

1.3 The finches of Jardín de Opuntia

The model organism for this study is the cactus finch (*Geospiza scandens*) and its probable hybrids. This small passerine bird belongs to the larger clade of the Darwin's finches and is endemic to the Galápagos Islands. Cactus finches feed mainly on giant Opuntia cacti (*Opuntia galapageia*) flowers, pulp and fruits (Grant and Grant 1980, Grant and Grant 1981, Millington and Grant 1983) and build dome-shaped nests between their big leaves (Addesso *et al*. 2020). Due to the high reliance on Opuntia cactus trees, the cactus finches are only found within cactus forests biomes. These habitats have been disappearing in recent times due to anthropogenic effects such as the introduction of alien invasive species (Addesso *et al*. 2020) and the transformation of natural areas into human settlements (Harvey *et al*. 2021). In the island of San Cristóbal, Jardín de Opuntia is the last remaining cactus forest.

Very little research has been done on the cactus finches of Jardín de Opuntia. Their population size has only been peripherally studied by general avian censuses of the area (Dvorak *et* al. 2020, Addesso *et al*. 2020). Local experts suggest that cactus finches of Jardín de Opuntia may be in danger due to the reduction of their habitat (p.c. Jaime Chaves). In contrast, their generalist cousin small ground finches (*Geospiza fuliginosa*) have seen an increase in population all across the archipelago, due to their adaptability to human-induced changes (Harvey *et al* 2021). This inequity in the two finches populations has lead local researchers to believe that they may be hybridizing (p.c. Jaime Chaves). The declining population of cactus finches might have prompted them to breach their reproductive isolation to seek a mate, potentially resulting in hybridization with the progressively more prevalent small ground finch.

1.4 Hypothesis

In order to assess hybridization within the two populations, this research will incorporate a novel technique for hybrids identification utilizing gut microbial data. To date, there has been no exploration into how hybridization affects the microbiome in any avian species. This gap in knowledge exists despite the extensive variety within recent avian species and microbiome surveys (Hird *et al*. 2015, Kropáčková *et al*. 2017).

To better understand the main hypothesis behind this study, a few notes from avian nurturing behaviour are necessary. In many altricial bird species, newly born offspring are fed through regurgitation (Groffen *et* al. 2008). In this phenomenon, one or both parents expel partially digested food from their mouth onto their hatchlings' open beaks. In cactus finches and all other Darwin's finches, this process is done by both parents as they alternatively leave the nest to forage for food (Boag and Grant 1984). Gut microbiome in regurgitating birds is inherited from the colonization of their guts derived from the parents' own microbiota present in the expelled digestate (Rosenberg *et al.* 2021). Based on this, if two parents from different species both feed their young through regurgitation, the microbiome of an hybrid offspring is hypothesized to show similarities in bacterial diversities and composition to both parental species.

To test this hypothesis, this research used fecal samples collected from the finches population in Jardín de Opuntia. Trough analyses of the microbiome DNA from the bird's excrements, cactus, small ground and presumed hybrid finches were all compared. Ultimately, this is done to identify if hybrids harbor a microbiome that resemble both of the parental species' ones.

2 Materials and methods

2.1 Study area and timeline

This research was carried out on Jardín de Opuntia (0° 56’ 18.92" S, 89° 32’ 54.93" W), on San Cristóbal island. The location is a partially secluded semi-natural cactus forest protected as part of the Parque Nacional Galápagos. The study site is a 1.4km by 0.12km area defined by a central 1.5km long path which terminates at a secluded beach on the east coast of the island. As the last cactus biome present on San Cristóbal, Jardín de Opuntia is the breeding ground of the only population of cactus finches on the whole island (Dvorak *et* al. 2020). Fieldwork was executed from January to April 2023, as these are the months which include the gradual start of the wet season when the birds are very active and population numbers are at their highest. Laboratory work was performed at the molecular biology laboratories of the GSC (Galápagos Science Centre) in Puerto Baquerizo-Moreno, San Cristóbal island. Lastly, data analysis on both morphological and microbiome data was performed using Python 3.12.

2.2 Catching the birds

Upon arrival at Jardín de Opuntia, the sampling location was chosen and a base camp for *in situ* analyses was set up nearby. Initial fieldwork involved capturing specimens at a fixed location for equipment and procedural acclimatization. Subsequently, sampling sites were randomly chosen across the entire path for broader coverage. Birds were captured using dense textured mist-nets, varying from 6m to 9m long, held up by poles 2m to 2.5m tall. Three to four nets were set up each morning, the number of nets and their dimensions were chosen based on the terrain, weather and number of people present.
The birds were mostly let passively fly into the nets. Occasionally, alarm calls were reproduced through phishing or recordings to attract the birds towards the nets.
Many different species of birds were caught in the nets but only Darwin’s finches were processed afterwards. Other avian species were instead immediately released (Appendix 1).

2.3 Field data collection

After being caught in the nets, the finches were relocated into cloth bird bags and transferred to camp. The birds were then taken out of the bag and ringed on their right leg. Some general characteristics were then assessed (Appendix 2) including species identification based on observations and quick comparisons to field guide information. The analysed birds were only finches and 6 species were present: small ground finch, medium ground finch (*Geospiza fortis*), cactus finch, warbler finch (*Certhidea fusca*), tree finch (*Camarhynchus parvulus*) and woodpecker finch (*Camarhynchus pallidus*).

Afterwards, a series of measurements were taken (Appendix 2) concentrating on the beak morphology. In more detail these were: beak length, beak depth and beak width (Figure 1). Measurements were done using electronic metal calipers (rounded to two decimal values over millimeters).

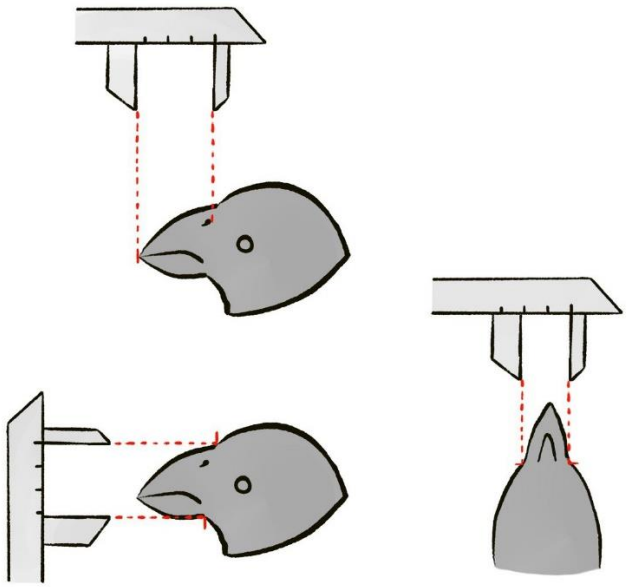


Figure 1: Methods of measuring beak morphology. On the top: beak length (calculated from nostril to point of the beak). On the bottom-left: beak depth (calculated from top to bottom of the beak). On the bottom-right (ventral view): beak width (calculated from side to side of the beak base).

While conducting the measurements, fecal samples were also collected. The birds were first placed in a feces collection apparatus (Knutie *et al*. 2018). This tool is crafted on site and rebuilt for every specimen. It consists in a paper lunch bag at the bottom of which a plastic dish is located, covered by a metal grid. The birds, as they are placed inside the bag, are naturally stimulated to defecate due to being released in a new environment. The metal grid helps the bird in roosting while the plastic dish helps keeping the surface sterile and easy for fecal sample collection. After 2 to 5 minutes the birds were taken out of the bag and the feces were pipetted inside a collection tube diluted with some distilled water. Afterwards, the tubes were conserved inside a cooler during the rest of the morning expedition and later brought to a -20°C freezer inside the laboratory.

2.4 Laboratory analyses

After data collection in the field, the samples were taken to the GSC's laboratory to be analyzed. The birds' fecal matter underwent different procedures to be translated into gut microbiome data. The laboratory studies were divided into 3 phases: extraction, amplification and sequencing.

The first step consisted in extracting DNA from the feces samples. First of all, the samples were selected and taken out of the fridge to be thawed by hand. Then, they were processed following PureLink's microbiome DNA purification kit protocol (Stool samples, Appendix 3). Few adjustments were implemented to maximize DNA yield. In particular, incubation times were increased from 8 minutes to 3 hours, while decreasing temperatures from 95°C to 65°C. Bead beating on the horizontal mixer was also increased from 15 minutes total to 1 hour. Afterwards, purified DNA was collected in different tubes and stored at -20°C.

During the second step, the extracted DNA was amplified through PCR (Appendix 4). This was done in preparation of the final sequencing step that required ulterior PCR amplification of DNA bound to barcode sequences. The availability of barcode sequences was very limited so the DNA was first trialed without barcodes in different concentrations to assess the optimal quantities and dilutions for amplification.

The third and final step was sequencing. This was done using Nanopore's MinION and its barcode libraries. The protocol 16s barcoding kit 1-24 (SQK-16S024) was followed (Appendix 5). Barcodes were implemented into each DNA sample though library preparation. The sequences were then amplified following the optimal values obtained from previous PCRs. Finally, the PCR results were pooled together and inserted into a Nanopore flow cell. The cell was then analyzed by Nanopore's MinION sequencer. DNA data was obtained as .fastq output after around 24 hours.

2.5 Data analysis

Three beak morphology measurements (length, width and depth) were analyzed as a base for the microbiome studies. A dataset was created containing the three measurements for each bird captured. K-means cluster analyses was performed on the dataset using Python's 'sklearn.cluster' as a machine learning model. To obtain the optimal number of clusters the elbow technique was utilized. The data was visualized as a 2D scatterplot using the 'plotly.express' package.
Lastly, to give statistic value to the clusters created using K-means, the groups were tested for significant differences. First a Shapiro-Wilk test was conducted on the beak morphology data to determine its normality. Consequently, a Kruskal-Wallis and post-hoc Dunn test were performed.

Afterwards the "hybridization visible though microbiome hypothesis" was tested through multiple analyses on Python. The .fastq sequencing data output from Nanopore's MinION was used for the microbiome analyses. Firstly, the data underwent a quality control using the NanoPlot program run on Python. Afterwards, it was controlled again, filtered and blasted against NCBI library using Nanopore's EPI2ME (protocol Fastq 16S 2023.04.21-1804452, split by barcode). The last filtering step was tweaked using the following parameters: BLAST-E quality value >7, sequence length between 1450bp and 1750bp, minimum coverage of 30%, minimum identity of 77%. The output data from EPI2ME contained taxonomic information for all blasted sequences at the genus level and was stored on a .csv file.
The latter file was then reorganized and visualized on bar graphs using the Python modules 'pandas' and 'plotly.express'. Three clustering techniques (K-means, DBScan and MeanShift) were performed on the reorganized microbiome dataset ('sklearn.cluster' model).
Multiple analysis were then conducted on the bacterial data using the clusters obtained from beak morphology. The microbial abundance of samples within each group was not normally distributed and very diverse (20 different total bacterial genera after grouping the 0.5% least common ones as "others"). Differences between the three morphology clusters were then assessed on the 4 most present bacterial genera, one by one. Differences between groups were tested using Mann-Whitney U and visualized with boxplots ('scipy.stats', 'itertools' and 'plotly.express' packages).
Simpson, Shannon and Chao alpha-diversity values (formulas in Appendix 6) were calculated for each sample. These indexes were then compared between the morphology clusters using Mann-Whitney U tests and visualized using boxplots ('NumPy' and 'plotly.express' packages).

3 Results

3.1 Morphology analyses

A total of 150 finches were caught in Jardín de Opuntia (Table 1). All of the processed birds belonged to the systematic group of the Darwin's finches. According to the initial field observations based on common recognizable morphological characteristics, 97 cactus finches and 20 small ground finches were sampled. Beak morphology variables (length, depth and width) were analyzed first in order to create a frame of reference for the microbiome analyses.

Table 1: Birds caught in Jardín de Opuntia. Last three columns illustrate average beak morphology values (in mm) with standard deviation error. Before dotted line: species identified with field book information. After dotted line: morphotypes obtained from k-means clustering. Extended version in appendix (Appendix 1).

Species	**N**	**Beak length**	**Beak depth**	**Beak width**
Darwin's finches				
Cactus finch	97	9.66±0.89	7.69±0.46	6.90±0.37
Small ground finch	20	8.64±0.69	7.74±0.72	6.86±0.53
Medium ground finch	18	12.97±0.82	12.65±1.14	10.49±1.07
Tree finch	11	8.40±0.66	7.46±0.61	6.78±0.49
Warbler finch	3	8.07±0.54	3.75±0.21	4.28±0.32
Woodpecker finch	1	10.81	9.23	7.16
TOTAL	**150**	**9.81±1.52**	**8.20±1.85**	**7.27±1.35**
Morphotypes				
Cactus finch	20	11.77±0.61	8.09±0.47	7.23±0.41
Hybrid	53	9.74±0.37	7.94±0.38	6.99±0.32
Small ground finch	53	8.87±0.37	7.42±0.36	6.67±0.35

Figure 2a is an overall look at all finches species in Jardín de Opuntia. The high variability in sizes is shown with some degrees of clustering for the most extremely sized species (warbler finch and medium ground finch) while cactus and small ground finches overlap in the middle section. The scatterplot on Figure 2b includes only the two species of interest (cactus and small ground finch) and illustrates the

closer morphological nature of the two birds. Further analyses only include cactus and small ground finches as those are the birds suspected to be hybridizing.

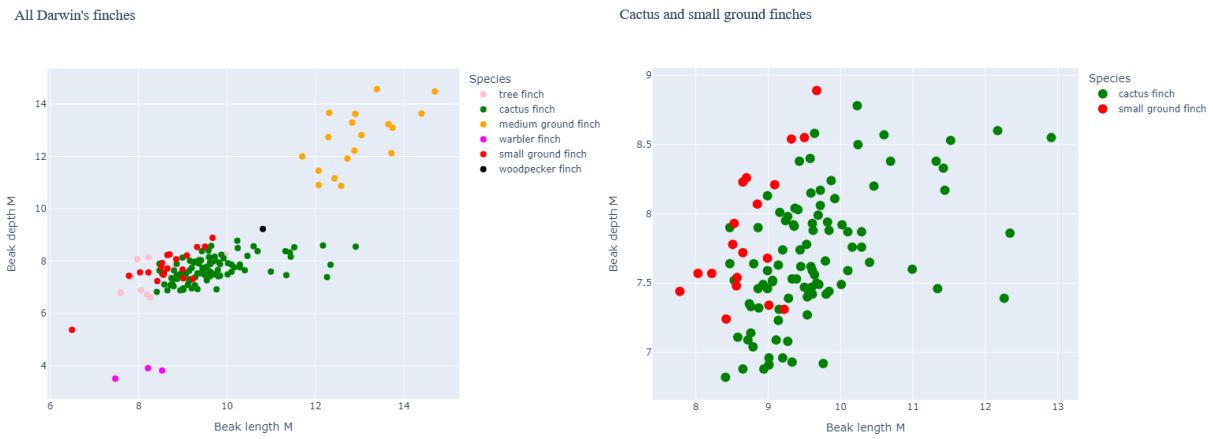


Figure 2: 2-dimensional scatterplots with axis defined by the three beak morphology measurements (x = length and y = depth, both in mm). 2a: on the left, all finches species are included. 2b: on the right, focus on cactus and small ground finches.

Beak length, beak depth and beak width were then utilized to optimally divide the data in clusters using k-means.

Before clustering, the elbow technique was implemented to reveal how the data is most accurately divided in 3 groups. The elbow technique in k-means clustering involves plotting the within-cluster sum of squares (WCSS) for different values of k (number of clusters). Then, it defines an optimal k by identifying the point where the rate of WCSS reduction slows down.

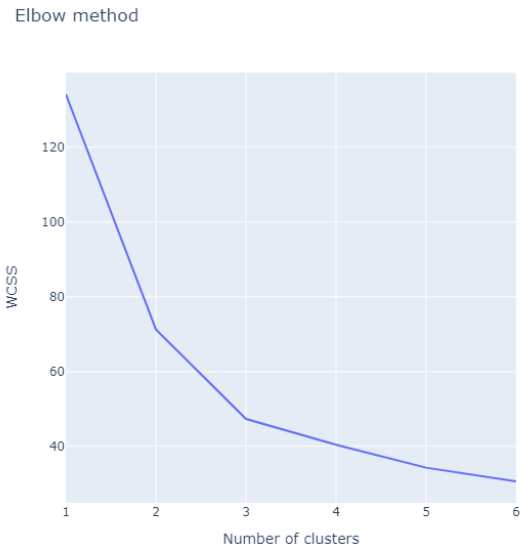


Figure 3: Elbow method for k-means clustering of the three beak dimensions between cactus ad small ground finches. X-axis is the number of proposed clusters (also called k). Y-axis is WCSS (Whitin Cluster Sum of Squares) which K-means clustering aims to minimize.

Figure 3 illustrates how the beak morphology data presented an optimal k = 3. Using this value, k-means was performed with 5000 iterations to ensure accurate clustering (Celebi *et al*. 2013).

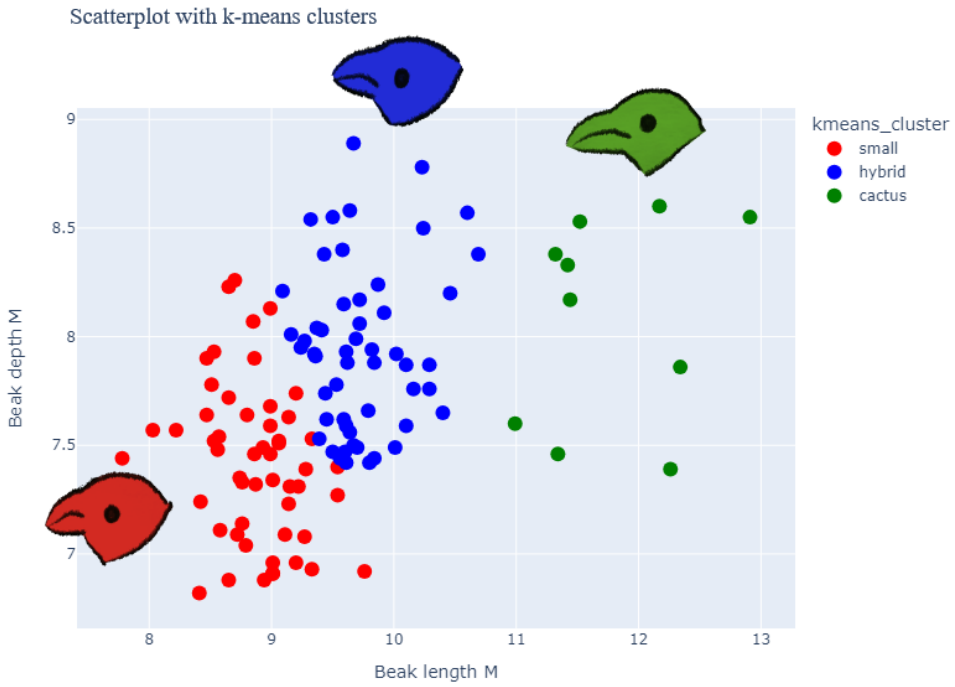


Figure 4: 2-dimensional scatterplot with beak length as the x-axis and beak depth as the y-axis (both in mm). G. fuliginosa and G. scandens individuals are grouped based on clustering values given by a k-means analysis. Illustrations of the three morphotypes are inserted for better visualization (no datapoints are being covered).

As shown in Figure 4, most individuals are part of the first two groups which represent the small ground finches and hybrid morphotypes, while the right-most group holds the generally longer beaked cactus finches. Based on these new findings, of the 117 birds caught: 54 were small ground finches, 20 were cactus finches and 53 were hybrids.

In order to give some more statistical backing to the k-means based clustering, the three groups were compared for similarities. The morphological data was not normally distributed (Shapiro-Wilk tests, p-value > 0.05), so a Kruskal-Wallis test was utilized. The clusters were found to be different (Kruskal-Wallis p-value < 0.01, Appendix 7). A Dunn post-hoc test revealed that for every beak morphology variable all groups were different from each other with the exception of two instances: cactus vs hybrid morphotypes considering beak depth (Dunn p-value = 0.51) and cactus vs hybrid morphotype considering beak width (Dunn p-value = 0.12).
The clusters obtained from k-means clustering were later used as a scaffold for comparisons on the finches' microbiomes.

3.2 Microbiome analyses

The initial hypothesis of this research is that an hybrid's microbiota resembles both parental microbiomes. In order to test this, bacterial abundance and diversity were analyzed and compared between putative hybrids and both parental species.

A total of 23 finches microbiota were successfully sequenced. Based on clustering from the beak morphology data: 6 were of small ground finches, 7 were cactus finches and 10 had a hybrid morphotype.

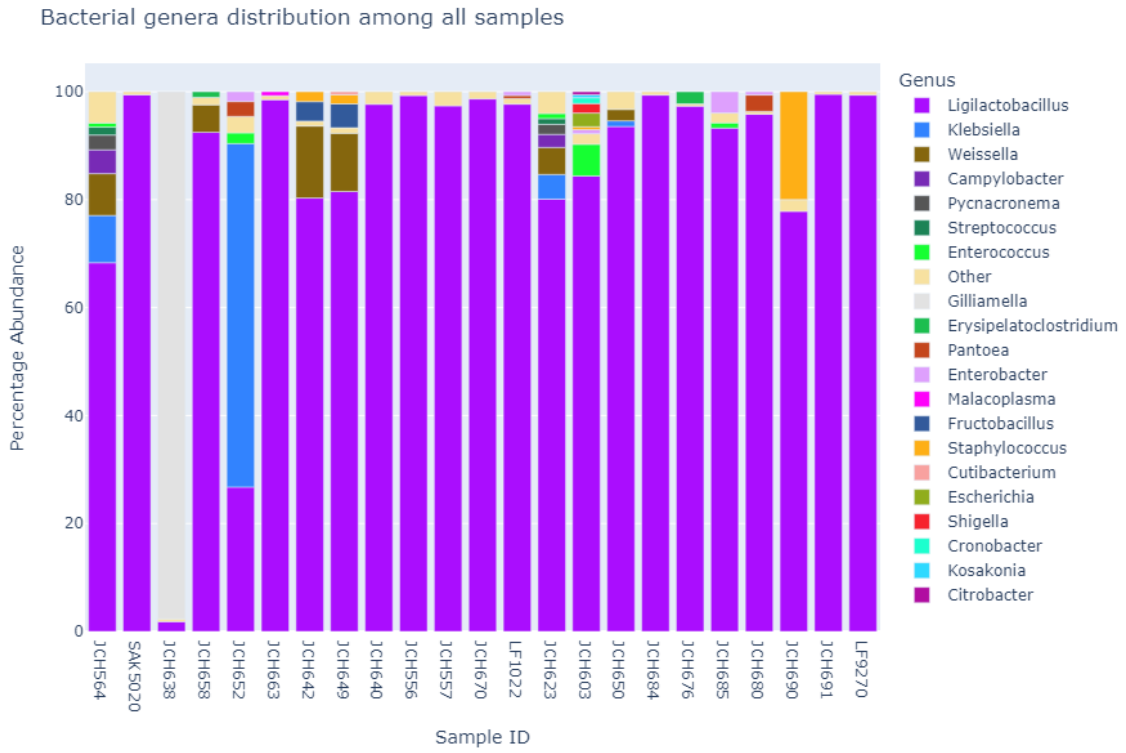


Figure 5: Bar plot illustrating bacterial diversity between the different samples on the x-axis. On the y-axis is the relative abundance of each bacterial genera.

Blasted data obtained from Nanopore's EPI2ME was first visualized using a bar plot (Figure 5). Scarcely ample genera were grouped together as "Other" when their individual relative abundance contribution within the sample was less than 0.5%.
The genus *Ligilactobacillus* was clearly the most dominant among most of finches' intestines with an abundance of 1,205,271 total hits (or 85.4% of the total bacteria). Other most present genera (which made up of more than 1% of the total bacteria) were: *Gilliamella* (76,440 hits, or 5.4%), *Klebsiella* (48,305 hits, 3.4%) and *Weisella* (30,140 hits, 2.1%).

Before using the morphological data as a base, different clustering techniques (k-means, DBSCAN, MeanShift) were used to find groups within all bacterial data. The results were not meaningful, leading to no clusters being reliably found (Appendix 8).
The large amount of different genera present in the samples and their sporadic distribution resulted in the data not being suited for multivariate analyses. The three clusters obtained from the beak

morphology analysis (small, hybrid and cactus) were then used as a base for the rest of the microbiome analyses.

In order to compare the bacterial data between the three groups, only the most common bacterial genera (with more than 10'000 hits) were considered individually. The data was not normally distributed (Shapiro-Wilk tests, p-value > 0.05). Mann-Whitney U tests were then employed to assess the similarity among groups in three distinct combinations (cactus vs hybrid, cactus vs small, hybrid vs small) for each bacteria. Almost all combinations confirmed that the three clusters were not dissimilar having a p-value well above 0.05 (Figure 6). Only the comparison for the genus *Gilliamella* between cactus and hybrid lead to the conclusion that the two groups were statistically different (Mann-Whitney U, p-value = 0.01). Also of note, though not below p-value, the comparison of *Gilliamella* abundance between hybrid and small resulted in a close rejection of the null hypothesis (Mann-Whitney U, p-value = 0.06).

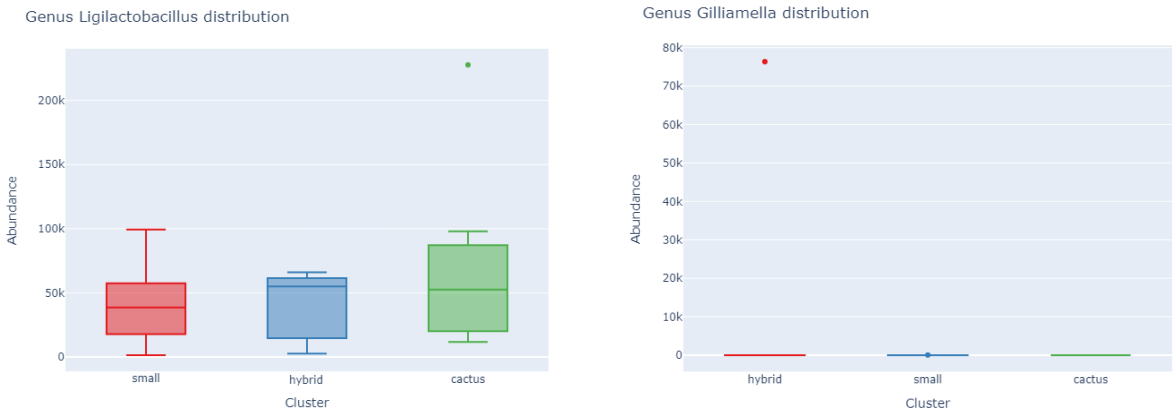


Figure 6: Box plots representing median distribution of bacterial genera on the three morphotypes. X-axis contains the different clusters. Y-axis is bacterial abundance values in number of total hits. On the left: genus Ligilactobacillus. On the right: genus Gilliamella. Extended graphs in Appendix 9.

To further investigate gut microbial differences between parental species and the hybrids, alpha-diversity indexes were calculated. Across all samples they revealed to be: Shannon 1.22, Simpson 0.71 and Chao 1949439.89 (Appendix 6).

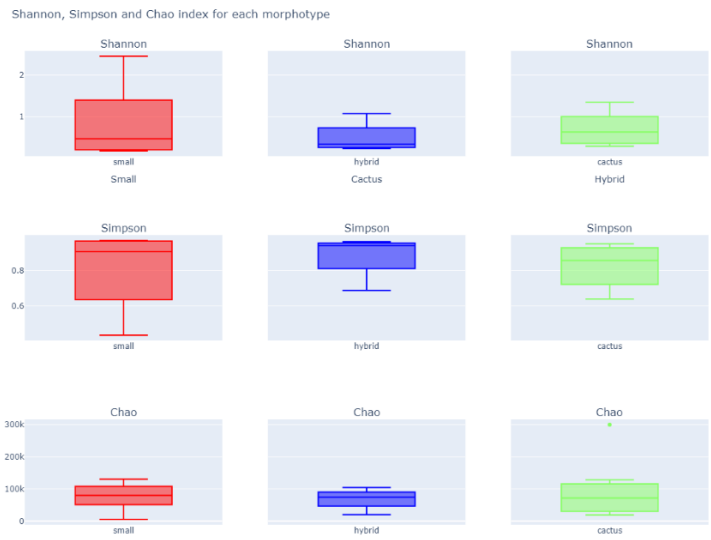


Figure 7: Alpha-diversity values of all samples, divided by morphotype. X-axis contains the three clusters. Y-axes represent the different values of each biodiversity index (in order: Shannon, Simpson, Chao).

The indexes were also obtained for each of the three morphotype groups (Figure 7). Then, for each alpha-diversity type, an ANOVA test was performed to find if there was any significant difference between cactus, small ground and hybrid morphotypes. No significant difference was found between different morphotypes alpha-diversity values (Shannon's p-value = 0.94, Simpson's p-value = 0.88, Chao's p-value = 0.57).

4 Discussion

4.1 Insights from morphology

Beak morphology data was analyzed as a base for microbiome studies, however it revealed some noteworthy insights on its own. The results from k-means clustering suggest a division in 3 groups rather than 2 based on beak morphology characteristics between cactus finches and small ground finches. As previously shown in Figure 4, these 3 groups are: "true" small ground finches and "true" cactus finches at the opposite extremities while the hybrids lie in-between. K-means clustering data conflicts partially with the original species identification from field book information. The 97 initial cactus finches were only 20 according to the morphology clusters, while the 20 small ground finches revealed to be 54 (Table 1). Interestingly, these numbers better reflect the initial expectations of the area, which were that the numbers of small ground finches were much higher than the cactus finches. To better determine the statistical value of the three groups, they were also analyzed statistically. The results of Kruskal-Wallis and post-hoc tests found that the clusters were all significantly different other than cactus finches and hybrids when considering beak depth and beak width. Regardless, the difference explained by beak length alone and by the three measurements together confidently confirms the presence of the three different morphotypes.

This data alone is unable to confidently assure the presence of a real functional 3rd middle species in between the two already existing, but should surely prompt further investigation on the topic. Until more detailed genetic studies are performed, the three groups can only be considered morphotypes and not different taxonomical units. Yet, the newly found third intermediate morphotype is likely result of hybridization since it is clearly composed of individual with middling morphological characteristics compared to the two defined extremes.

4.2 The microbiome-based approach

The initial hypothesis was: if an offspring is fed by both parents through regurgitation, then an hybrid will inherit a microbiome which resembles both of the parental species' ones.

Using the three morphotypes as groups, the microbial data was analyzed looking for similarities and differences between the parental species and the hybrids. Comparisons considering only the most present bacterial genera revealed the high similarity between all three clusters. Only in the case of the genus *Gilliamella* deviant results were obtained. Variations in *Gilliamella*'s abundance suggest a difference between cactus finches and hybrids (null hypothesis of similarity is rejected). Due to the very low sample size (23 total birds sequenced), this outlier alone can hardly suggest a real difference between the sampled groups. In addition, group comparison between alpha-diversity values also revealed the high level of similarity of the three clusters.

Overall the analyses all point towards cactus finches, small ground finches and hybrids having too similar gut microbiota for any difference to be detected. Interestingly, the hypothesis is still supported since the hybrid individuals did resemble the parental species in microbiome composition. Regardless, this revealed to be uninformative since cactus and small ground finches themselves did not differ from one another. The novel microbiome-based approach of hybrid identification on avian species then remains untested. One of the first point of improvement for future research is then to improve the hypothesis to better reflect the importance of the parental species having different gut microbiotas.

The way of inheritance of gut microbiota may be one of the reasons why no difference was found between the two parental and the hybrid's microbial compositions. The bird's intestines are firstly colonized by bacteria in the immediate post-natal environment, but during juvenile growth, changes in the microbiota can still take place (Miller *et al* 2021). The influence of diet, environment and species on microbiome have all been identified in several vertebrates. In mammals, a study on 60 different species (Ley *et* al. 2008) and another on bats (Carrillo-Araujo *et* al. 2015) both highlighted the impact of host phylogeny and feeding strategies on microbial composition. In avian investigations, variation in the microbiota is often related to bird's species (Hird *et al*. 2015, Kropáčková *et al*. 2017) but also to diet (Waite *et al*. 2014) and locality (Garcia-Amado *et al*. 2018). In this thesis only the inheritance from parental strains was tested. The three morphotypes live in sympatry and may share a similar diet. The influence of environmental and feeding derived microorganisms may have overruled the original differences between cactus and small ground finches microbiomes by homogenizing their diversity. Ecological surveys of the finches populations in Jardín de Opuntia have yet to be performed. The reliance on research from other islands (on Daphne Major, Grant and Grant, 1981) may have exaggerated the differences in diet between finches in the area of interest.

Another point of improvement is taxonomic accuracy. Nanopore's MinION is a reliable machine that enables research in very remote environments but its accuracy is not yet on par with other high end sequencers (Laver *et al*. 2015). Because of this, the genetic data was only blasted at the genus level. The high abundance of *Ligilactobacillus* (85.4% of all bacteria) if seen at the species level may reveal differences in the data that at a higher taxonomic level are not perceivable. A novel paper inquiring on microbiome's role in avian speciation did not find significant difference between two closely related passerine species while inquiring at the genus level (Sottas *et al*. 2021). A study on mouse hybrids (Wang *et al*. 2015) instead found differences in the hybrid microbiome when looking at the bacterial species level. This last research on mice was done utilizing a 454 GS-FLX sequencer which has lower error rate than Nanopore's MinION. Further studies should then aim to utilize more accurate and less error prone sequencing techniques to be able to inquire the data at a finer taxonomic level.

In addition, a better base of comparison should be made utilizing genetic data from the bird tissues. This assessment of hybrids and their degree of hybridization is most effective when considering that individuals often fall on a spectrum between the genetic characteristics of the pure parental species (Ottenburghs 2018). This variation arises from occasional back-crossing of the hybrids. Gene flow and introgression can only be identified when considering the organisms' own genome rather than the variably inherited microbiome. Generations of hybrid mating may have diluted the differences between the birds microbiota (Huxel 1999), so investigation on the birds' genotypes will reveal more detailed information on the hybridization status of the finches in Jardín de Opuntia.

4.3 Conclusion

The hypothesis for this study was supported but did not lead to a usable microbiome-based hybrid identification approach. The biological notions at the base of the hypothesis should then be re-investigated with a novel, more detailed hypothesis which includes the necessity for differences between parental microbiome compositions. This research should be considered as a pioneer study on a technique, species and area that have all yet to be completely understood. Further investigations should inquire the finches of Jardín de Opuntia utilizing more accurate sequencing techniques, comparison with the host's own genetic data, and primary assessment of their dietary preferences.

5 Bibliography

Addesso, A. M., Harvey, J. A., Vaziri, G. J., Verrett, T. B., Albert, L., Arthur, C., ... & Knutie, S. A. (2020). Effect of introduced parasites on the survival and microbiota of nestling cactus finches (Geospiza scandens) in the Galápagos Islands. *Journal of Ornithology*, *161*, 1011-1019.

Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press.

Boag, P. T., & Grant, P. R. (1984). Darwin's finches (Geospiza) on Isla Daphne Major, Galápagos: breeding and feeding ecology in a climatically variable environment. Ecological Monographs, 54(4), 463-489.

Carrillo-Araujo, M., Taş, N., Alcantara-Hernandez, R. J., Gaona, O., Schondube, J. E., Medellin, R. A., ... & Falcon, L. I. (2015). Phyllostomid bat microbiome composition is associated to host phylogeny and feeding strategies. *Frontiers in microbiology*, *6*, 447.

Celebi, M. E., Kingravi, H. A., & Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert systems with applications*, *40*(1), 200-210.

Cullen, C. M., Aneja, K. K., Beyhan, S., Cho, C. E., Woloszynek, S., Convertino, M., ... & Rosen, G. L. (2020). Emerging priorities for microbiome research. *Frontiers in microbiology*, *11*, 136.

Dvorak, M., Fessl, B., Nemeth, E. R. W. I. N., Anchundia, D., Cotin, J., Schulze, C. H., ... & 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.

García-Amado, M. A., Shin, H., Sanz, V., Lentino, M., Martínez, L. M., Contreras, M., ... & Domínguez-Bello, M. G. (2018). Comparison of gizzard and intestinal microbiota of wild neotropical birds. *PLoS One*, *13*(3), e0194857.

Grant, B. R., & Grant, P. R. (1981). Exploitation of Opuntia cactus by birds on the Galápagos. *Oecologia*, *49*, 179-187.

Grant, P. R., & Grant, B. R. (1980). Annual variation in finch numbers, foraging and food supply on Isla Daphne Major, Galápagos. *Oecologia*, *46*, 55-62.

Groffen, H., Watson, R., Hammer, S., & Raidal, S. R. (2008). Analysis of growth rate variables and postfeeding regurgitation in hand-reared Spix's macaw (Cyanopsitta spixii) chicks. *Journal of avian medicine and surgery*, *22*(3), 189-198.

Hammer, T. J., Sanders, J. G., & Fierer, N. (2019). Not all animals need a microbiome. *FEMS microbiology letters*, *366*(10), fnz117.

Harvey, J. A., Chernicky, K., Simons, S. R., Verrett, T. B., Chaves, J. A., & Knutie, S. A. (2021). Urban living influences the nesting success of Darwin's finches in the Galápagos Islands. *Ecology and Evolution*, *11*(10), 5038-5048.

Helbig, A. J., Salomon, M., Bensch, S., & Seibold, I. (2001). Male-biased gene flow across an avian hybrid zone: evidence from mitochondrial and microsatellite DNA. *Journal of Evolutionary Biology*, *14*(2), 277-287.

Henry, L. P., Bruijning, M., Forsberg, S. K., & Ayroles, J. F. (2021). The microbiome extends host evolutionary potential. *Nature communications*, *12*(1), 5141.

Hird, S. M., Sánchez, C., Carstens, B. C., & Brumfield, R. T. (2015). Comparative gut microbiota of 59 neotropical bird species. *Frontiers in microbiology*, 1403.

Huxel, G. R. (1999). Rapid displacement of native species by invasive species: effects of hybridization. *Biological conservation*, *89*(2), 143-152.

Knutie, S. A., & Gotanda, K. M. (2018). A non-invasive method to collect fecal samples from wild birds for microbiome studies. Microbial ecology, 76, 851-855.

Kong, S., & Kubatko, L. S. (2021). Comparative performance of popular methods for hybrid detection using genomic data. *Systematic Biology*, *70*(5), 891-907.

Kropáčková, L., Těšický, M., Albrecht, T., Kubovčiak, J., Čížková, D., Tomášek, O., ... & Kreisinger, J. (2017). Codiversification of gastrointestinal microbiota and phylogeny in passerines is not explained by ecological divergence. *Molecular Ecology*, *26*(19), 5292-5304.

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science*, *359*(6372), 224-228.

Laver, T., Harrison, J., O'neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the oxford nanopore technologies minion. *Biomolecular detection and quantification*, *3*, 1-8.

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., ... & Gordon, J. I. (2008). Evolution of mammals and their gut microbes. *science*, *320*(5883), 1647-1651.

Li, Z., Wright, A. D. G., Si, H., Wang, X., Qian, W., Zhang, Z., & Li, G. (2016). Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses. *Environmental microbiology reports*, *8*(6), 1016-1023.

Miller, A. K., Westlake, C. S., Cross, K. L., Leigh, B. A., & Bordenstein, S. R. (2021). The microbiome impacts host hybridization and speciation. *PLoS biology*, *19*(10), e3001417.

Millington, S. J., & Grant, P. R. (1983). Feeding ecology and territoriality of the cactus finch Geospiza scandens on Isla Daphne Major, Galápagos. *Oecologia*, *58*, 76-83.

Morais, L. H., Schreiber IV, H. L., & Mazmanian, S. K. (2021). The gut microbiota–brain axis in behaviour and brain disorders. *Nature Reviews Microbiology*, *19*(4), 241-255.

Ottenburghs, J. (2018). Exploring the hybrid speciation continuum in birds. *Ecology and evolution*, *8*(24), 13027-13034.

Ottenburghs, J. (2023). How common is hybridization in birds?. *Journal of Ornithology*, 1-8.

Randler, C. (2004). Frequency of bird hybrids: does detectability make all the difference?. *Journal of Ornithology*, *145*, 123-128.

Rosenberg, E., & Zilber-Rosenberg, I. (2021). Reconstitution and transmission of gut microbiomes and their genes between generations. Microorganisms, 10(1), 70.

Short, L. L. (1969). Taxonomic aspects of avian hybridization. *The Auk*, *86*(1), 84-105.

Sommer, F., & Bäckhed, F. (2013). The gut microbiota—masters of host development and physiology. *Nature reviews microbiology*, *11*(4), 227-238.

Sottas, C., Schmiedová, L., Kreisinger, J., Albrecht, T., Reif, J., Osiejuk, T. S., & Reifová, R. (2021). Gut microbiota in two recently diverged passerine species: evaluating the effects of species identity, habitat use and geographic distance. *BMC Ecology and Evolution*, *21*(1), 1-14.

Waite, D. W., & Taylor, M. W. (2014). Characterizing the avian gut microbiota: membership, driving influences, and potential function. *Frontiers in microbiology*, *5*, 223.

Wang, J., Kalyan, S., Steck, N., Turner, L. M., Harr, B., Künzel, S., ... & Baines, J. F. (2015). Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nature communications*, *6*(1), 6440.

6 Appendix

Appendix 1- Extended Table 1

Table A1: Birds caught in Jardín de Opuntia. Last three columns illustrate average beak morphology values (in mm) with standard deviation error. Before first dotted line: species identified with field book information. Between first and second dotted line: morphotypes obtained from K-means clustering. After second dotted line: non-Darwin's finches avian species found in the location of interest.

Species	N caught	Beak length	Beak depth	Beak width
Darwin's finches				
Cactus finch	97	9.66±0.89	7.69±0.46	6.90±0.37
Small ground finch	20	8.64±0.69	7.74±0.72	6.86±0.53
Medium ground finch	18	12.97±0.82	12.65±1.14	10.49±1.07
Tree finch	11	8.40±0.66	7.46±0.61	6.78±0.49
Warbler finch	3	8.07±0.54	3.75±0.21	4.28±0.32
Woodpecker finch	1	10.81	9.23	7.16
TOTAL	**150**	**9.81±1.52**	**8.20±1.85**	**7.27±1.35**
*Morphotypes**				
Cactus finch	20	11.77±0.61	8.09±0.47	7.23±0.41
Hybrid	53	9.74±0.37	7.94±0.38	6.99±0.32
Small ground finch	53	8.87±0.37	7.42±0.36	6.67±0.35
*Other birds***				
Yellow warbler				
Galápagos flycatcher				
Galápagos mockingbird				

*Based on beak morphology k-means clustering

**Freed immediately, not counted

Appendix 2- A series of other characters were identified on initial field processing: sex, age, presence of brood patch and cloacal protuberance, fat levels, past pox and molting.

Also blood was collected and other morphological characters were calculated: Head width, length and depth, tarsus length, wing chord and mass.

Here are collected some unorganized notes on the different values and measurements assessed, they are not part of the main research body since they didn't fit the scope of the research or were considered not significant.

Sex was only distinguishable in few species: small ground finch, cactus finch and medium ground finch. In these species, the males had a darker plumage while the females had a lighter brownish one. Age was determined based on beak colour and presence of meaty parts around the beak, the birds would then be identified as juvenile or adults. The plumage colour was also different in males and could be a further indicator of their age. It was scored from 1 to 5, where 5 was the darkest and corresponded to the oldest age. These were the same methods used by Darwin to score age in his studies.

Brood patch and cloacal protuberance (BP and CP) are two signs of reproductive age in finches, they were scored as present or absent. BP is defined by a bald featherless spot in the bird's belly, usually but not exclusively present in female birds to aid in heat transfer to the offspring during hatching. CP is a ring shaped protuberance surrounding the cloaca in male birds that consists of a collection of seminal fluids, usually present in preparation for reproductive events.
Fat levels were scored as absent, low or high. They are identified by the deposition of white/yellow fat tissue in the neck, especially around the furcula.
Past pox infection leave very recognizable signs in finches. These consist in hindlimbs malformations and subsequent loss of digits or talons. It was scored taking into consideration the specific location of the digits missing.
Moulting could be present or not and if present could be located in multiple parts of the body. As such it was scored as: head, body, tail or wing. If located on the wing It was also specified if symmetrical or not on both wings.

The finches were bled restraining them with one hand and puncturing them using a thin needle with the other. The puncture was made on the wide spread wing, at the level of the brachial vein. After penetration the blood was collected using a small capillary tube and later transferred over an FTA blood collection card. Information about the ID, location, species and sex of the birds were written on the card and buffer was sprayed over it for better conservation.

Afterwards, a series of measurements were taken concentrating on the beak morphology. In more detail these were: beak length, beak depth, beak width, head length, head depth, head width, tarsus length, wing chord and mass. The latter was done on a small electric scale (rounded to one decimal value after grams) with the bird located inside a small cardboard tube to restrain it. Beak, head and tarsus measurements were taken using electronic metal calipers (rounded to two decimal values over millimeters). The wing chord was calculated using a small metal ruler (rounded to the millimeter).

Typically, these collections were performed prior to the weighing step, which was intentionally saved as the final stage due to the increased risk of the bird flying away.

Appendix 3- PureLink's Stool samples Microbiome 16s DNA extraction protocol.

MAN0014266_PureLinkMicrobiome_Stool_UG.pdf

Available on: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0014266_PureLinkMicrobiome_Stool_UG.pdf

Appendix 4- PCR details

•Denaturation phase 1: 1 minute at 95°C, 1 repetition

•Denaturation phase 2: 20 minutes at 95°C, 30 repetitions

•Annealing phase 3: 30 minutes at 55°C, 30 repetitions

•Annealing phase 4: 2 minutes at 65°C, back to phase 2, 30 repetitions

•Extension Phase 5: 5 minutes at 65°C, 1 repetition

•Extension Phase 6: Until the end at 4°C, 1 repetition

Appendix 5- Nanopore's 16s barcoding kit 1-24 (SQK-16S024) protocol.

Available on: https://community.nanoporetech.com/docs/prepare/library_prep_protocols/16S-barcoding-1-24/v/16s_9086_v1_revy_14aug2019

Appendix 6- details of alpha-diversity values equations.

Shannon:

$$H = -\sum i = 1Spi\mathbf{ln}(pi)$$

Where:
- H = is the Shannon Diversity Index.
- pi =proportion of individuals of species i relative to the total number of individuals.

Chao:

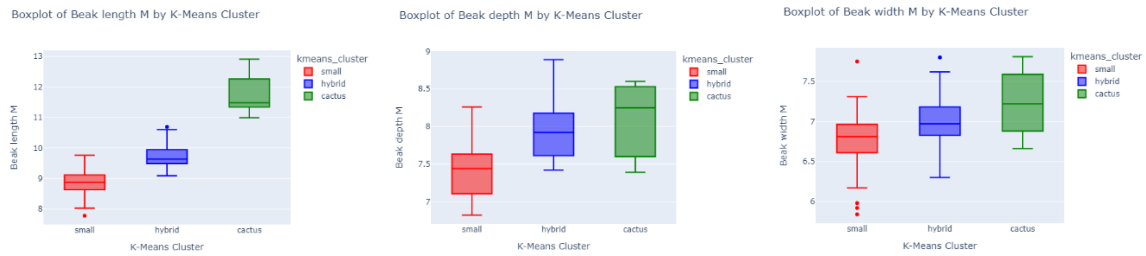
Chao1 = S+(F(F−1)/ 2(M1/M2))
Where:
- S = observed number of species.
- F = number of singletons (species represented by a single individual).
- M1 = number of species with only one individual.
- M2 = number of species with only two individuals.

Simpson:

$$D = 1 - \sum i = 1S(Nni)2$$

Where:
- D = Simpson Diversity Index.
- ni = number of individuals of species i
- N = total number of individuals in the community.

Appendix 7- Kruskal-Wallis and Dunn post-hoc in boxplots.



Figures A1: Hereby the results from Kruskal-Wallis and Dunn post-hoc are visualized for each beak morphology value, divided by cluster. On the second and third boxplots, hybrid and cactus are not significantly different.

Appendix 8- Matrix and other beta-diversity clustering attempts.

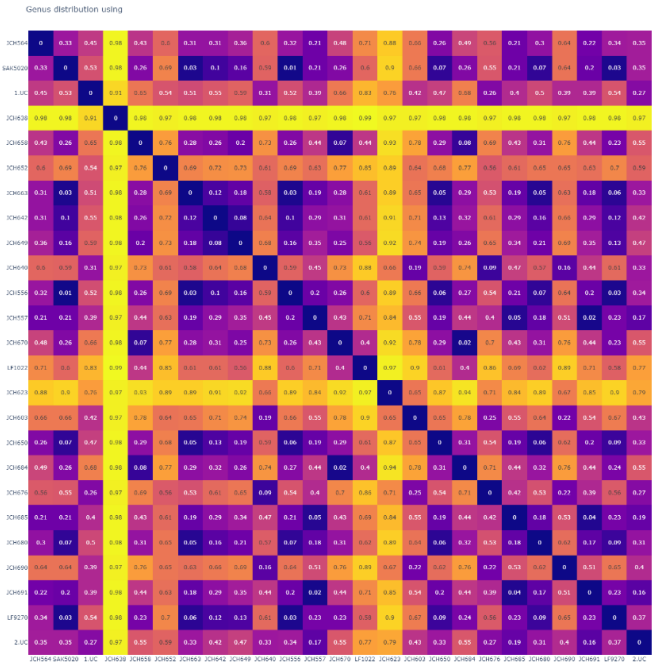
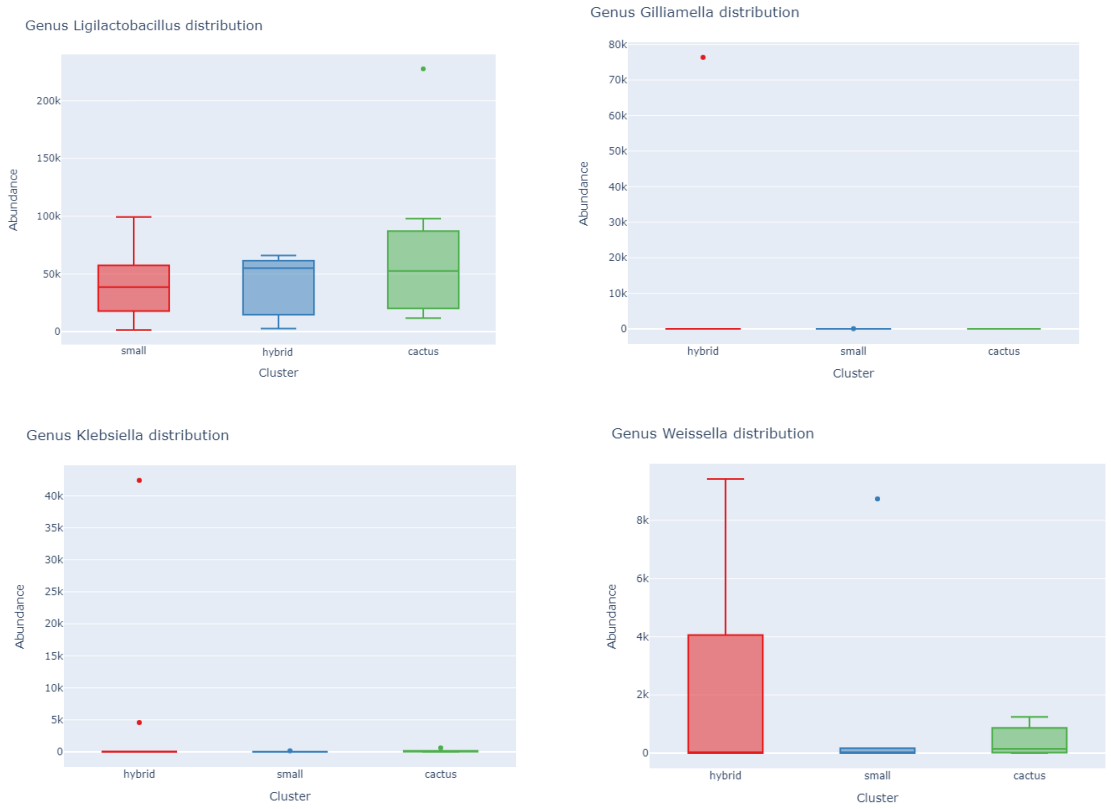


Figure A2: Dissimilarity matrix based on beta-diversity values considering one on one comparison between all samples. X and Y axes both include a copy of all sequenced samples.

A matrix of bray Curtis beta-diversity values was created between all samples to alternatively visualize their little differences. Most samples have values closer to 0 representing a high degree of similarity, with

the exclusion of the outliers JCH638 and JCH62. Data was attempted to be clustered using 4 different clustering techniques, but they all failed to provide meaningful results.

Appendix 9- Other bacterial groups comparison graphs.



Figures A3: Box plots representing median distribution of bacterial genera on the three morphotypes. X-axis contains the different clusters. Y-axis is bacterial abundance values in number of total hits. On the top-left: genus Ligilactobacillus. On the top-right: genus Gilliamella. On the bottom-left: genus Klebsiella. On the bottom-right: genus Weissella.