

Nearly (?) sterile avian egg in a passerine bird

Martin Těšický^{1,2,3,6,*}, Lucie Schmiedová^{1,2}, Tereza Krajzingrová¹, Mercedes Gomez Samblas^{1,4}, Petra Bauerová⁵, Jakub Kreisinger¹, Michal Vinkler¹

¹Faculty of Science, Department of Zoology, Charles University, Viničná 7, 128 43 Prague, Czech Republic

²Institute of Vertebrate Biology, v.v.i., The Czech Academy of Sciences, Květná 8, Brno 603 65, Czech Republic

³Institute of Paleontology, Domestication Research and History of Veterinary Medicine, Ludwig Maximilian University of Munich, Kaulbachstr. 37 III, 80539 Munich, Germany

⁴Faculty of Science, Department of Parasitology, Campus Universitario de Fuentenueva, University of Granada, Profesor Adolfo Rancano, 18071 Granada, Spain

⁵Division of Air Quality, Czech Hydrometeorological Institute, Tušimice Observatory, Tušimice 6, 432 01 Kadaň, Czech Republic

⁶Present addresses: Institute of Vertebrate Biology, v.v.i., The Czech Academy of Sciences, Květná 8, Brno 603 65, Czech Republic and Institute of Paleontology, Domestication Research and History of Veterinary Medicine, Ludwig Maximilian University of Munich, Kaulbachstr. 37 III, 80539 Munich, Germany

*Corresponding author. Faculty of Science, Department of Zoology, Charles University, Viničná 7, 128 43 Prague, Czech Republic, EU. Tel: +420221951845; Fax: +420221951841; E-mail: martin.tesicky@natur.cuni.cz

Editor: [Leluo Guan]

Abstract

During early ontogeny, microbiome affects development of the gastrointestinal tract, immunity, and survival in vertebrates. Bird eggs are thought to be (1) initially sterile (*sterile egg hypothesis*) and (2) colonized after oviposition through *horizontal trans-shell migration*, or (3) initially seeded with bacteria by *vertical transfer* from mother oviduct. To date, however, little empirical data illuminate the contribution of these mechanisms to gut microbiota formation in avian embryos. We investigated microbiome of the egg content (day 0; E0-egg), embryonic gut at day 13 (E13) and female faeces in a free-living passerine, the great tit (*Parus major*), using a methodologically advanced procedure combining 16S rRNA gene sequencing and microbe-specific qPCR assays. Our metabarcoding revealed that the avian egg is (nearly) sterile, but acquires a slightly richer microbiome during the embryonic development. Of the three potentially pathogenic bacteria targeted by qPCR, only *Dietzia* was found in E0-egg (yet also in negative controls), E13 gut and female samples, which might indicate possible *vertical transfer*. Unlike in poultry, we have shown that major bacterial colonization of the gut in passerines does not occur before hatching. We emphasize that protocols that carefully check for environmental contamination are critical in studies with low-bacterial biomass samples.

Keywords: egg microbiome; embryo; gastrointestinal tract microbiota; passerine bird; pathogenic bacteria; sterile egg

Introduction

Gut microbiota plays a paramount role in host physiology, affecting nutrient digestion (Bäckhed et al. 2005), gastrointestinal tract (GIT), and immune system regulation (Ost and Round 2018), gut-brain-axis signalling (Strandwitz 2018), and even onset of diseases (Honda and Littman 2012). One of the most puzzling and debated question in microbiology is whether the gut microbiota is formed before or after birth/hatching in animals, yet this is methodologically challenging to test. While early next-generation sequencing (NGS) studies suggested that transmission of the maternal microbiome to embryos before birth may be universal in animals, including humans (e.g. Funkhouser and Bordenstein 2013), these studies are now thought to have suffered from increased environmental contamination and sequencing artefacts (Eisenhofer et al. 2019, Kennedy et al. 2023). It is now assumed that human placenta and foetus are sterile under physiological conditions and that the neonatal gut is colonized only after birth (Walker et al. 2017, de Goffau et al. 2019, but see Kennedy et al. 2023). Although some recent studies evidence the existence scarce and low-abundant microbial communities in foetal gut (Rackaityte et al. 2020, Bi et al. 2021, Mishra et al. 2021), these results were questioned in the recent community review by Kennedy et al. (2023), as they likely also suffered from contamination during the sampling procedure. Is a

bird egg also initially sterile or do females deposit any bacteria into the egg to direct the initial embryonic microbiome development? This is a relevant question, since early bacterial colonizers can shape GIT and the immune system development, influencing survival and composition of chick microbial communities after hatching (Hansen et al. 2015, Roto et al. 2016). However, the sources of egg microbial colonization and timing of embryonic microbiota development still remain poorly understood in birds.

Three mutually nonexclusive hypotheses have been proposed to explain the origin of the initial gut microbiota in birds. *Sterile egg hypothesis* (1) assumes that the bird egg is initially formed sterile in the female reproductive tract (Roto et al. 2016). This would require either the absence of bacteria in the female oviduct or presence of host oviduct physiological filters, preventing bacterial colonization of the developing eggs (Lee et al. 2019). The *sterile egg hypothesis* was particularly influential in the era of culture-based studies (reviewed in Roto et al. 2016), but more recently some egg microbial communities have been detected in eggs by 16S rRNA gene metabarcoding (see below). Avian eggs are well protected by several physical (cuticle, crystalline eggshell, and shell membranes; Liong et al. 1997, Lunam and Ruiz 2000, D'Alba and Shawkey 2015) and chemical mechanisms (e.g. antimicrobial peptides with bacteriolytic activity; Mann 2007, Gantois et al. 2009, Cuperus et al.

Received 14 February 2023; revised 2 December 2023; accepted 18 December 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of FEMS. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

2013) that create a unhostile environment for invading microbes. Hence, major microbial colonization of GIT could occur only after hatching. Nevertheless, as an alternative it has been proposed that despite the egg protection, bacteria can still colonize the eggs (2) through *vertical transfer* while being formed in the maternal reproductive tract and/or (3) through *horizontal trans-shell migration* from the environment after egg laying (Pedroso 2009, Roto et al. 2016).

For *vertical transfer*, bacteria are assumed to pass to the developing eggs mainly from the female oviduct, originating either from passively ascending colonies in cloaca, or being actively transported by macrophages or dendritic cells directly from the female GIT (Gantois et al. 2009). *Vertical transfer* has been suggested particularly for pathogens, such as *Salmonella*, where experimentally orally infected hens laid contaminated eggs (Keller et al. 1995, Gantois et al. 2009, Pedroso 2009). However, the frequency of such vertical transmission is typically very low (e.g. for *Salmonella* detected in egg whites ranging usually between 0% and 4.5% but occasionally reaching up to 20%, depending on the study; reviewed in Gantois et al. 2009). A similar low frequency of vertical transfer has also recently been suggested for commensal microbiota in domestic chickens (Lee et al. 2019) and passerines (Trevelline et al. 2018) using NGS metabarcoding.

Bacterial horizontal trans-shell migration assumes that bacteria colonize eggs only postlaying through the eggshell pores (Bruce and Drysdale 1994). Thus, the embryonic gut microbial communities would be established by bacteria migrating from the nest environment (Van Veelen et al. 2018, Lee et al. 2019), combining the female microbiome with the microbiome on the nest material, generally referred to as *nidobiome* (Campos-Cerda and Bohannan 2020). Some bacteria, such as environmental *Neisseria*, have been documented to penetrate the eggshells in the greater white-fronted goose eggs (*Anser albifrons*; Hansen et al. 2015) and cause embryonic mortality. Compared to eggs, chicken embryos, show diversified microbial communities as revealed in embryonic gut both by classical microscopy (Kizerwetter-Świda and Binek 2008) and more recently by 16S rRNA gene metabarcoding (e.g. Ding et al. 2017, 2022, Lee et al. 2019). These are dominated by taxa such as *Pseudomonas*, *Janthinobacterium*, *Acinetobacter*, *Stenotrophomonas*, and *Meganomonas*, which could primarily originate from eggshell, GIT or cloaca of females (Lee et al. 2019).

The level of bacterial egg colonization may vary between species, e.g. between free-living and captive species or species with precocial and altricial developmental modes. However, some caution is needed when interpreting the results of different NGS studies. For example, not all NGS studies adequately integrated negative controls of various types (extraction, amplification, and so on.) or performed amplifications in PCR replicates. In contrast to the chicken studies (e.g. Ding et al. 2017, 2022, Lee et al. 2019), Grond et al. (2017) revealed only negligible microbiota (not significantly different from the negative controls) in embryonic GIT of two Arctic shorebirds. This raises the question of whether the observed marked differences in microbiota composition of egg content and embryonic GIT in birds are due to differences between species, or whether the results could also be biased by methodological differences. Therefore, the extent to which the different colonization mechanisms contribute to the establishment of the embryonic microbiota in wild birds is not yet resolved.

In our study, we simultaneously investigated in wild birds whether the egg is initially sterile (i.e. testing the *sterile egg hypothesis*) and how bacterial colonization occurs (if detected) during embryonic development (*vertical transfer* vs. *horizontal bacterial*

trans-shell migration). We employed an innovative approach to analyze the microbial profiles of egg content and embryonic GIT in the great tit (*Parus major*). During the 2018 breeding season, we collected a total of 240 microbial samples from 57 nests of a free-living great tit population breeding in Prague, Czech Republic. We adopted a methodologically improved 16S rRNA gene metabarcoding approach along with specific qPCR assays to reveal sample contaminants masking the natural variation in microbial composition. Our objectives were (i) to examine the initial microbiota of egg contents shortly after laying (embryonic day 0, E0) to determine whether the eggs are initially sterile, and (ii) to compare the initial egg-content microbiota with embryonic GIT microbiota just before hatching (embryonic day 13, E13) and breeding adult female (maternal) faecal microbiota to assess the different colonization mechanisms. Particular attention was paid to pathogenic bacteria with potential negative effects on host fitness. To validate our ability to detect bacteria later in E13, we also experimentally administered *Enterococcus faecium* inoculum to E0 eggs that were followed alive to E13.

Materials and methods

Sampling design and sample collection

The sampling was conducted in a free-living great tit population breeding in artificial nest boxes in a deciduous forest at the edge of Prague, Czech Republic, EU (50°08'12.4"N, 14°27'57.2"E; see Těšický et al. 2021, 2022 for more details on the study site) during their breeding period in April and May 2018. In total, we collected 52 eggs to sample microbiome of the egg content (E0-egg), 118 eggs to sample E13 embryonic gut (unmanipulated, E13-nat, as well as *Enterococcus*-manipulated eggs, E13-Ent, see below), and 34 maternal female faecal samples (see Table S2 in Supporting Information 1 (SI1) for number of samples in different categories). Overall, these samples represented 58 nests, of which the complete sample set was available from 18 nests. The time of breeding was determined by regular inspections of the nest boxes (about 2–7-day intervals, adjusted to the estimated hatching date). We numbered all eggs in the clutches based on their laying order with a permanent marker.

To describe natural microbiota composition in E0-eggs, we collected one freshly laid egg per clutch ($N = 52$) within 1 day after laying and transporting it to the laboratory. There, in a laminar biosafety cabinet (Jouan MSC 12, ThermoFisher Scientific, Carlsbad, USA) its surface was cleaned with 96% ethanol and DNA remover to prevent contamination. Subsequently, the entire egg content, i.e. ~200–300 μ l, was aspirated aseptically with an insulin syringe (B Braun, catalogue number 9151125, Melsungen, Germany) after puncturing the eggshell. Then samples were stored deep-frozen in PCR-clean cryotubes (Simport, Canada) at -80°C .

To describe natural microbiota composition in embryonic GIT (E13-nat), we collected either one or two E13 eggs per clutch ($N = 66$). Two E13 eggs per nest were collected from a total of 25 clutches to assess whether microbiota is more similar within a nest than between nests. To confirm that putative bacteria present in E0-egg can be reliably detected using our methods, we injected within 2 days after laying a subset of the eggs with *E. faecium* (reference strain: NCIB 11181; probiotics Lactiferm Basic 5, Chr. Hansen, Hørsholm, Denmark, catalogue number L-0265). Specifically, we injected 10 μ l of *E. faecium* at a concentration of either 10^7 (high dose) or 10^4 (low dose) colony-forming units into the first egg per clutch (as equal results were obtained, these

were categorized collectively as E13-Ent, $N = 56$, see the section 'Results') and 10 μl of PBS (Sigma Aldrich, catalogue number D5652-50 L; $N = 53$) into a second egg, which served as a control (E13-PBS). After manipulation, treated eggs were sealed at the injection site with a superglue, returned to their nests and incubated together with E13-nat until E13 when all E13 eggs were collected (see SMMO 3 for details on the *in ovo* application, Fig. 1 for a timeline and sample design scheme and Fig. 2 for the methodological overview). The collected E13 eggs were transported to the laboratory and kept in an incubator (Brinsea Octagon 20 Advance Incubator, Brinsea Products Inc, Titusville, USA) at constant temperature and humidity (37.5°C and 60%) until their aseptic dissection (maximum 6 h after collection). In the biosafety cabinet, the embryos were aseptically removed from the eggs, and placed on sterile Petri dishes (ThermoFisher Scientific, catalogue number 1011RR). Then the embryos were decapitated and their gut (the GIT part between the gizzard and cloaca) were taken and stored frozen at -80°C in RNA later (Qiagen, catalogue number 76106, Hilden, Germany). Tissue samples were successfully collected from a total of 23 E13-Ent and 29 E13-PBS eggs, i.e. 41.1% and 54.7% of the original number, respectively, due to embryonic mortality.

To describe maternal microbiota, we also noninvasively collected faecal samples from adult females ($N = 34$) using our previously described methodology (Kropáčková et al. 2017a). Briefly, birds were captured in mist nets when the nestlings were 7–14 days old. Immediately after the capture, they were placed into fresh paper bags for ca. 15–20 min. Then faecal samples were collected with sterile microbial swabs (minitip FLOQSwabs, Copan, Italy) and stored in RNA later at -80°C . We also measured tarsus length and weight (for later calculation of standardized body mass as the ratio of tarsus length and weight) and collected blood and feather samples (not included in this study). The age of the birds was assessed based on differences in plumage colouration of primary and secondary coverts (Svensson and Baker 1992) and from ringing records. Finally, we ringed all birds with an aluminium ring with a unique code of the Czech Bird Ringing Centre, National Museum in Prague. The research was carried out under the applicable laws of the Czech Republic and the European Union. The experiment was approved by the Environmental Protection Department of the Prague City Hall (permit number S-MHMP-1061728/2010/OPP-V-790/R-235/Bu) and the Ministry of the Environment of the Czech Republic (permit number 22003/ENV/16-1009/630/16). Permission for capturing and ringing of adult birds was granted by the Czech Bird Ringing Centre of the National Museum in Prague.

DNA isolation

To maximize the efficiency of our microbial DNA extraction with minimal contamination risks, we first performed a comparison of five bacterial DNA extraction kits (for details see [Supplementary Material and Methods Online 1 \(SMMO 1\)](#) in [SI1](#)). Finally, all microbial DNA samples were extracted using only the DNeasy PowerSoil Kit (Qiagen, catalogue number 47016) with some modifications (see [SMMO 2](#) in [SI1](#)) in a laminar biosafety cabinet. Samples were homogenized using a vortex with horizontal adapter (catalogue number 13000-V1-24; MO BIO Laboratories, Inc, Carlsbad, USA) for 10 min at maximum speed to optimize DNA isolation efficiency, and extracted DNA was eluted to 55 μl with an elution buffer. As starting material we used: (1) for E0-eggs 200 μl of homogenized egg content mixed with 200 μl of sterile water (to prevent pellet formation); (2) for E13 embryos the intestinal samples,

and (3) for adult females whole faecal samples. To avoid cross-contamination between different biological sample types, each sample type was extracted separately, strictly following the principles of clean molecular work (decontamination procedures and manipulations minimizing the risk of between-sample contamination). We also included isolation negative controls (INCs) with nuclease-free water (the same batch as used for egg content dilution) which were processed separately for each sample type in the following counts: for egg content samples ($N = 7$), embryonic samples ($N = 16$), and female faecal samples ($N = 3$).

Microbial metabarcoding

Our metabarcoding approach was based on amplification of the V3–V4 region of the 16S rRNA gene. Two different protocols were used for both egg content and embryonic samples to maximize the probability of detecting the microbiota varying in the primer-targeted sequences and minimize the impact of primers and amplification kit selection. These issues could be particularly important in low-bacterial biomass studies. In protocol 1 (P1), no-chloroplast amplifying primers 335F (CADACTCCTACGGGAGGC) and 769R (ATCCTGTTTGMTMCCCVCRC) (Dorn-In et al. 2015) together with KAPA2G Robust PCR Kit (Kapa Biosystems, catalogue number KK5005, Wilmington, USA) allowed more specific ASV amplification. In protocol 2 (P2), using the universal bacterial primers S-D-Bact-0341-b-S17 (CCTACGGGNGGCWGCAG) and S-D-Bact-0785-a-A-21 (GACTACHVGGGTATCTAATCC) (Klindworth et al. 2013) together with KAPA HIFI Hot Start Ready Mix (Kapa Biosystems, catalogue number 07958935001) with proofreading activity allowed amplification of greater ASV diversity. Both primers amplify partially overlapping regions. All samples were amplified with a combination of both protocols, except for female faecal samples that were run only with the no-chloroplast amplifying primers due to high proportion of dietary-derived chloroplast sequences (Kropáčková et al. 2017a). Both primer sets were tagged with 10 bp oligonucleotides for multiplexing Nextera™ DNA Sample Prep Kit (Illumina®-Compatible, catalogue number GA09115, San Diego, USA). For each sample and primer set, PCR was performed in technical duplicates to check the consistency of microbial profiles (see [Table S1](#) in [SI1](#) for details on the PCR conditions). Different biological sample types were amplified in different plates to prevent cross-contamination. Two negative PCR controls per plate (hereafter referred to as no PCR template controls, NTCs; RNA free water; catalogue number 760011596, Qiagen) were used. Then all PCR products were run on 1.5% agarose gel and the PCR product concentrations were assessed based on gel band intensity using GENOSOFT software (VWR International, Belgium). Samples were pooled into several pools based on their concentration and were purified with SPRIselect paramagnetic beads (Beckman Coulter Life Sciences, USA). To remove PCR nonspecificities, PCR products in the range of 520–720 bp were excised by Pipin Prep instrument using 1.5% Agarose Cassettes, dye-free, int. Standards (Pippin Prep, 250 bp–1.5 kb, catalogue number 341CDF1503, Biozym, Hessisch Oldendorf, Germany). Subsequently, the concentration of purified pools was checked by Qubit Fluorometer (ThermoFisher Scientific) with Qubit dsDNA BR Assay Kit (ThermoFisher Scientific, catalogue number Q32850) and pooled in equimolar concentrations. Finally, the resulting amplicon libraries were sequenced using MiSeq Illumina platform with 2×300 bp paired-end reads and v3 chemistry (Illumina) at the Central European Institute of Technology (CEITEC, Brno, Czech Republic). E0-egg and embryonic GIT samples (i.e. the low-bacterial biomass samples) were sequenced in

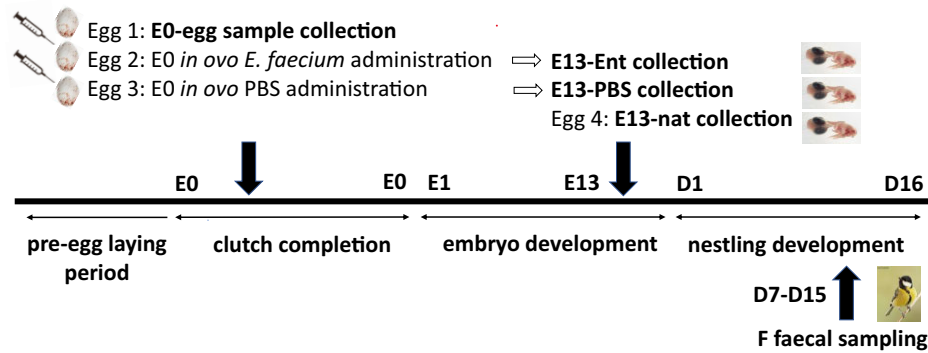


Figure 1. Timeline and experimental set-up of the great tit study. Generally, five microbiota samples were collected per nest: E0-egg—egg content sampled on embryonic day 0 (E0), E13-nat—E13 intestinal sample from a nonmanipulated egg, E13-Ent—E13 intestinal sample from an *Enterococcus*-treated egg, E13-PBS—E13 intestinal sample from a control, PBS-injected egg, and F—female faecal samples collected between days 7 and 15 (D1–D15) of nestling age. E and D above the axis indicate the day of embryonic and chick development, respectively.

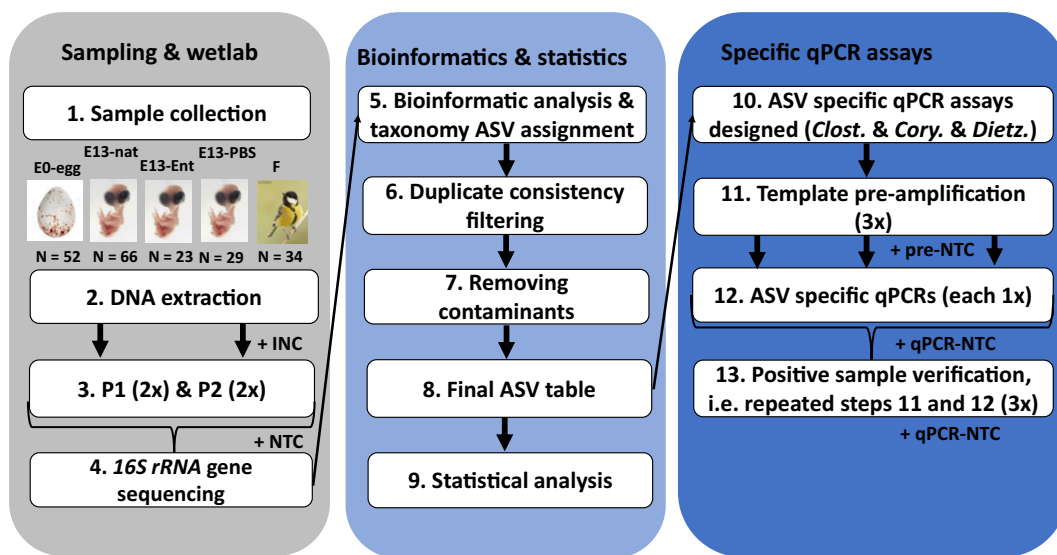


Figure 2. Schematic representation of the procedure for determining microbial profile in egg content, embryonic intestine, and female faecal samples in the great tit study. (1) Different types of biological samples were collected in the field: E0-egg—egg content sample at embryonic day 0, E13-nat—E13 intestinal sample from nonmanipulated egg, E13-Ent—E13 intestinal sample from *Enterococcus*-treated egg, E13-PBS—E13 intestinal sample from a control PBS-injected egg, and F—adult female faecal sample. (2) Bacterial DNA was extracted including isolation negative controls (INC). We applied (3) two protocols for 16S rRNA gene DNA microbial genotyping: protocol 1, P1 with no-chloroplast amplifying primers and protocol 2, P2 with universal bacterial primers and 16S rRNA amplicon libraries were prepared. The libraries were sequenced (4) followed by (5) bioinformatic analysis and taxonomic assignment of bacteria. (6) Only amplicon sequence variants (ASVs) that were consistently present in both technical duplicates were retained. (7) We removed potential contaminants and (8) based on the abundance of particular ASVs in different biological sample types, we further (9) statistically analyzed sequencing data. Based on these results, (10) we developed specific probe-based qPCR assays to detect potentially pathogenic bacteria (*Clostridium*, *Dietzia*, and *Corynebacterium*). (11) To increase the sensitivity of qPCR assays, we preamplified the template DNA with bacterial universal primers including negative control of preamplification (pre-NTC). (12) ASV-specific qPCRs were then performed with these preamplified DNA templates and negative control of qPCR (qPCR-NTC). (13) Positive samples from qPCR were further verified by independent amplification (repeated steps 11 and 12). Numbers in brackets correspond to the number of technical replicates per sample.

different sequencing runs than the female faecal samples (i.e. the high-bacterial biomass samples) because DNA concentration differed significantly between these biological sample types and faecal samples could have cross-contaminated other samples (see Table S1 in Supporting Information 2 (SI2) for the metadata).

Bioinformatic processing of the sequence data and identification of microbial taxa

Samples were first demultiplexed and primers were trimmed by skewer software (Jiang et al. 2014). Using dada2 (Callahan et al. 2016), we filtered out low-quality sequences (expected number of errors per read less than 2), denoised the quality-filtered fastq

files and created an abundance matrix representing read counts for each amplicon sequence variant (ASV) in each sample. Using uchime (Edgar et al. 2011) and gold.fna database (available at <https://drive5.com/uchime/gold.fna>), we identified chimeric sequences and removed them from the abundance matrix. We aligned the ASV sequences obtained with the two different primer sets (P1 and P2; Table S1 in SI1), which amplify slightly different 16S rRNA gene regions, using DECIPHER R package (Wright 2015) and retained only the region that overlapped between amplicons. The mean amplicon sequence length before trimming was 413 bp (median = 416 bp, min = 389, max = 439) for P1 and 425 bp (median = 427 bp, min = 352, max = 450) for P2, and 404 bp (median = 411 bp, min = 352 bp, max = 416 bp) after trimming. The same

sequences after this trimming step were considered as one ASV. ASVs were taxonomically assigned to the lowest distinguishable taxonomic level using the RDP classifier (80% confidence threshold, Wang et al. 2007) and SILVA SSU reference database (v. 138; released December 2019; Quast et al. 2013). From all downstream analyses, we excluded all ASVs that were classified as 'Chloroplast', 'Mitochondria', 'Eukaryota', or not assigned to any bacterial phylum. We also removed all samples with a low number of sequences (< 50 reads with sequencing artefacts or low number of reads insufficient for statistical analysis).

For the different sample types and protocols, we assessed the consistency of microbial profiles between technical duplicates using Procrustean analysis based on Bray–Curtis dissimilarities calculated based on ASVs proportions in each sample (Kreisinger et al. 2017). Depending on the protocol used, E0 eggs, INC or NTC had lower consistency of technical PCR duplicates than female and embryonic samples (Table S6 in S11). We then removed any ASVs that were not consistently present in both technical duplicates for a given sample (putative contamination and sequencing and PCR artefacts). From the samples without duplicates, we excluded all ASVs that were present in one sample and not in any other duplicated sample. Read counts for duplicated samples were merged for all later analyses.

Importantly, we used the Decontam package (Davis et al. 2018) to identify and subsequently eliminate putative contaminating ASVs whose prevalence was elevated in INC and NTC samples compared to all great tit samples and/or that were more prevalent in samples with a low concentration of metagenomic DNA (as measured by PCR product concentration). The analysis was carried out separately for egg and embryonic samples and for female faecal samples. Finally, we compared these filtered metabarcoding results to the lists of the most commonly contaminating ASVs compiled by Salter et al. (2014), Eisenhofer et al. (2019), and Stinson et al. (2019).

Statistical analysis of the sequence data

Variation in ASV richness (i.e. number of ASVs detected in each sample) between sample types and the PCR protocols was compared after excluding *Ralstonia* and *Enterococcus* (putative contaminants not identified by Decontam, see below) using generalized linear mixed effect models (GLMM) with negative binomial distribution and sample identity as a random effect (R package glmmTMB). Because community dissimilarity cannot be calculated for sample pairs that do not contain bacterial reads (which was relatively common for egg and embryo samples), we could not perform standard beta diversity analyses (e.g. PCoA ordination or PERMANOVA analyses). In addition, we performed a differential abundance analysis examining the variation in abundance of individual ASVs between experimental groups and PCR protocols. These analyses were based on mixed models with negative binomial distribution, where the number of reads for each ASV in each sample served as the response variable, whereas the experimental group and PCR protocol were considered predictors. To account for differences in sequencing depth between samples, we also included log-transformed read counts (increased by one) for each sample at the offset. To improve model convergence, differential abundance analyses were performed for a subset of ASVs detected in at least 5% of the samples. To avoid false positives due to multiple testing, we calculated false discovery rate (FDR; Benjamini and Hochberg 1995) based on the resulting probability values and considered only those results as significant where the FDR was less than 0.05.

Using the microeco package, we created Venn diagrams (Liu et al. 2021) showing overlaps of ASVs and genera between different biological sample types and PCR protocols. To test whether the female faecal microbiota is more similar to the microbiota of the own egg or embryo than expected by chance, we compared the differences in Jaccard and Bray–Curtis dissimilarities between the female faecal microbiota and the microbiota of the own vs. foreign egg (E0-egg)/embryo (E13-nat) using the Wilcoxon test. This analysis was only performed for PI and for the dataset with *Ralstonia* and *Enterococcus* ($N_{E0-egg} = 47$, $N_{E13-nat} = 66$, $N_F = 30$), as there were not enough samples available for statistical analysis after its exclusion. We also used the Wilcoxon test to examine whether embryos from the same nest are more similar in their microbial composition than embryos from different nests. Jaccard dissimilarities were calculated after rarefaction of the abundance matrix ($N = 51$ sequences per sample, i.e. the minimum number of reads in the dataset). Bray–Curtis dissimilarities were calculated based on relative ASV abundances. All statistical analyses were performed using R software, v.4.1.1 (R Core Team 2017).

Quantitative PCR screening of potentially pathogenic bacteria

Based on the metabarcoding results, we developed ASV-specific probe-based 16S rRNA gene (DNA) qPCR assays to detect potentially pathogenic bacteria in egg contents and embryos. We set the following criteria for selecting target ASVs: (1) the ASV presence in E0-egg as well as in E13-nat samples, (2) the ASV absence from any INCs and NTCs, (3) the ASV being not previously reported as a contaminant of laboratory plastics or chemicals (based on a literature survey), and finally (4) the ASV being known as a potential pathogen of the gastrointestinal or urogenital tract in animals (pathogens have a greater impact on host physiology and fitness). In cases where the selected ASVs were similar in sequence to other ASVs found in INCs or NTCs, we designed the qPCR primers and probes to target only a more dissimilar ASV variant that could not be amplified nonspecifically together with any potential contaminant. Following these criteria, we designed three qPCR assays: (i) for *Corynebacterium* (Barbosa and Palacios 2009, Risely et al. 2018), (ii) for *Clostridium* (Tsiodras et al. 2008, Benskin et al. 2009), and (iii) for *Dietzia* (Koerner et al. 2009, Olowookere et al. 2022), see Table S3 in S11 for primer and probe sequences.

To increase the sensitivity of the qPCR assays, all samples were preamplified for 30 cycles with bacterial universal primers (Klindworth et al. 2013; the same as in P1) but with high fidelity and accurate polymerase Platinum SuperFi PCR I Master Mix (ThermoFisher Scientific, catalogue number 12351), see Table S5 in S11 for more details. To minimize DNA binding to the plastic surface, we used low-binding plastics and added 1 µl of 0.1 ng/µl tRNA carrier (i.e. spike water, Qiagen, catalogue number 1068337) to each tube with stock eluted DNA. Preamplification was performed in technical triplicates and different biological sample types were run on different plates to minimize the risk of cross-contamination. Three negative preamplification controls (NTC-pre-amp; i.e. spike water) were included on each plate. Preamplified PCR products were then diluted 3x with spike water and used as a template for ASV-specific qPCR (see Table S5 in S11 for PCR conditions).

ASV-specific qPCRs were performed with Luna Universal Probe qPCR Master Mix (New England Biolabs, Inc., catalogue number E3006, Ipswich, Massachusetts, USA) under the manufacturer's specified conditions (Table S5 in S11) using the Light Cycler 480 (Roche Applied Science) in a 384-well plate format (Roche Applied

Science, catalogue number 04729749001). All assays were performed in technical triplicates, with the preamplified DNA from the technical triplicate used as a template for independent monoplex qPCR (allowed by the high repeatability between technical triplicates in qPCR). DNA sequence standards (IDT, gBlocks Gene Fragments; Table S4 in SI1 in serial dilutions of 10^9 – 10^1 were used to estimate the qPCR efficiency ($E = 1.995$ for *Clostridium*, $E = 1.989$ for *Corynebacterium*, and $E = 1.835$ for *Dietzia*). In addition, each plate also contained three template-free negative controls (spike water; NTC-qPCR) and three NTC-pre-amp (as mentioned above).

For each triplicate, we calculated the mean C_p values using the second derivative method (second derivative Max) implemented in LightCycler 480 SW 1.5 (Roche Applied Science). A replicate was considered positive when $C_p < 36$, which corresponds to ~ 1 – 10 DNA molecules after the preamplification (determined using the gBlock standard curves). However, given the expected high stochasticity of PCR amplification at very low template concentrations, all E0-egg and embryonic GIT samples where at least one of the triplicates was positive (C_p values < 36) were reassessed with another independent preamplification and qPCR (a total of six technical replicates per sample were obtained). Finally, we defined samples as positive only if the C_p values were in at least 2/6 replicates < 36 and simultaneously, the positivity was confirmed in both independent qPCR runs (see Fig. 2 for the procedure overview). Due to the preamplification step, we could only interpret qPCR results as semiquantitative by determining the number of replicates in the samples reaching the amount of bacterial DNA over ~ 1 – 10 molecules per reaction (estimated based on gBlocks standards). The original measured qPCR data are in Table S2 in SI2.

Using the qPCR data, we statistically tested whether *Corynebacterium*, *Clostridium*, and *Dietzia* occur more frequently in biological samples than in negative controls (INC, NTC-qPCR, and NTC-pre-amp). We applied generalized linear models (GLMs) with quasi-binomial distribution where the ratio of the number of positive replicates ($C_p < 36$) to the total number of replicates was a dependent variable and biological sample type was an independent variable. Separate models were built for each combination of the assay (i.e. given ASV) and biological sample type (here E0-nat, E13-nat, and F), except in cases where the given ASV was not confirmed in any sample by independent qPCR (e.g. for *Clostridium* or *Corynebacterium* in E0-nat and E13-nat; see all models M1-5 in SI3). Plots were generated using ggplot2 (Wickham 2016), boot (Canty and Ripley 2021), and ggeffects packages (Lüdtke 2018). All statistical analyses were performed using software R, v. 4.1.1 (R Core Team 2017).

Results

Microbial data filtering and general information

From the total 857 samples sequenced (including biological duplicates), we obtained sequences for 855 samples, obtaining 1996 026 reads and identifying in total 1382 ASVs. Applying various bioinformatic filtering steps and eliminating effects of contamination significantly reduced the total number of samples and ASV diversity included in the later analysis. This pattern was particularly evident in the low-bacterial biomass samples. For effects of the individual filtering steps in different sample types see Figure S2 and Tables S2A–S2E in SI1. First, we eliminated all samples with very low numbers of sequences obtained (< 50 reads), reducing the representation of samples in our analysis by 9.59% ($n = 773$) and 0.01% of all reads. Second, we removed inconsistent ASVs between

duplicates, sharply reducing the numbers of ASVs (to 152 out of 1370 ASVs). This step had only a weak impact on the numbers of reads used for the analysis (decline by 9.81%), indicating the presence of many low-abundant ASVs that were inconsistently represented in the duplicates of the full dataset. Third, for each biological sample we combined the filtered sequences from the duplicates, retaining 416 samples for further analysis.

Additional attention has been further paid to specific sample contaminations. Using Decontam package in R we statistically removed 12 potentially contaminating ASVs in embryonic and egg content samples and two in female faecal samples (see Figure S1 and Table S7 in SI1 for full taxonomy), accounting for 4.84% drop in the read counts [see Figure S2 and Table S2 in SI1 for all filtering steps]. Especially the low-bacterial biomass samples were dominated by the genus *Ralstonia*, which comprised in total 30.1% of the reads. It was also the only ASVs distributed across all sample types (Figure S9A in SI1 and Table S3A in SI2). This finding and the fact that *Ralstonia* is known as a common contaminant of plastics and solutions (Ryan and Adley 2014) suggest that our samples were contaminated with this bacterium although it was not detected in the Decontam analysis. Furthermore, despite all our efforts to keep our procedure clean, *E. faecium* used for the *in ovo* treatment might have cross-contaminated some of the other embryonic samples (particularly E13-PBS; Figures S4, S5 in SI1). Therefore, in all subsequent analyses we present the results conservatively with and without *Ralstonia* and *Enterococcus*. Filtering out *Ralstonia* and *Enterococcus* substantially reduced the number of reads, from 1 711 376 to 401 089 reads. In the final dataset, a total of 128 ASVs remained, occurring in only 191 samples containing sequences, of which only 1236 reads (0.06% of the original number of reads in the full dataset before filtering) belonged to E0-egg, 45 488 reads (2.28%) to E13-nat and 272 337 reads (13.64%) to female faecal samples. This indicates that despite the increased number of PCR cycles in E0-egg and E13-nat samples, both E0-egg and E13-nat samples contained only very few bacterial sequences.

Our results on microbiota composition obtained using the two metabarcoding protocols were generally very consistent. Since the P1 data contain fewer sequencing artefacts, we primarily show the results of the P1 approach here and provide the P2 results and the results with *Ralstonia* and *Enterococcus* in SI for comparison. However, in cases where a comparison of data is necessary or where there are discrepancies between the P1 and P2 results, we also highlight this in the main text.

Microbial profiles in egg content, embryos, and females

The experimental groups differed significantly in ASV richness (GLMM: Δ D.F. = 4, $\chi^2 = 70.021$, $P < .0001$). Higher ASV richness was found using the P2 than P1 protocol (GLMM: estimate [\pm S.E.] = 0.30187 ± 0.08769 , Δ D.F. = 1, $\chi^2 = 11.989$, $P < .0001$, Fig. 3). According to Tukey *post hoc* comparisons, female faecal samples had significantly higher ASV richness compared to the other categories, except E13-nat. E0-egg and E13-Ent had significantly lower richness compared to all other categories except E13-PBS, whose richness was not different from E13-nat. According to differential abundance analyses, three ASVs (from the genera *Methylobacterium*, *Mycobacterium*, and *Shingomonas*) were significantly more abundant in the E13-nat group and a single ASV (from the family Xanthobacteraceae) was more abundant in the E12-PBS group (Figure S1 and Figure S3 in SI1). There were no ASVs whose abundance varied significantly between PCR protocols.

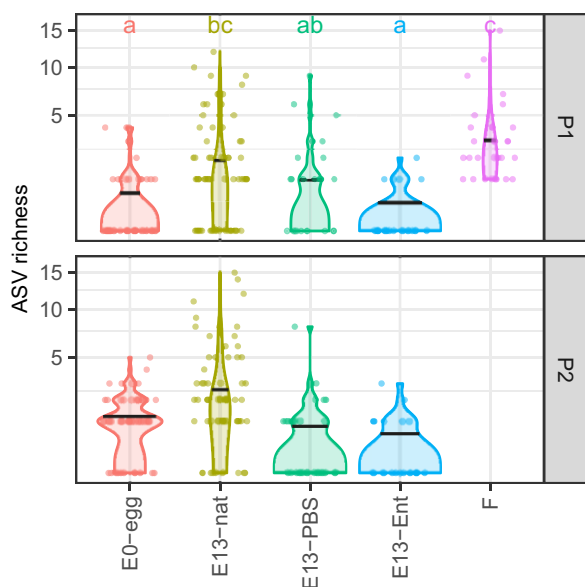


Figure 3. Comparison of amplicon sequence variant (ASV) richness between biological sample types using violin plot in the great tit. Results for two PCR protocols are shown (P1 and P2, see the section ‘Methods’ for more details). The black horizontal lines within the violin plots indicate the median values. Lower case letters in the upper part of the plot indicate significant differences between groups according to Tukey post hoc tests. If the letters for two groups are different, the groups differ significantly in their mean ASV richness ($P < .05$). For visual purposes only, the y-axis is squared root scaled.

In eggs (E0-egg), after the filtering and exclusion of *Ralstonia* out of 52 samples only eight samples had enough sequences (mean = 154.5 reads per sample), while majority (84.6%) of the samples did not contain any biologically reliable bacterial sequences (data obtained using P1; Table S2A–S2E and Figures S4, S5 in SI1). Overall, only 11 ASVs were detected with mean 2.38 ASV per sample (Tables S2D, S2E in SI1). Among the most abundant ASVs in the E0-egg samples, the family Xanthobacteraceae and genera *Cutibacterium* and *Mycobacterium* were detected in at least five samples and *Spingomonas* was detected in two samples (Table 1; Fig. 4; Figures S4–S7 in SI1). *Clostridium* was revealed in one E0-egg sample. P2 revealed similar ASV numbers as P1 (15) but detected two putative pathogens, *Dietzia* (one sample) and *Corynebacterium* (three samples).

In embryonic GIT (E13-nat), out of 66 samples only 31 had enough sequences after all the filtering steps (mean = 1467.4 reads per sample), while 53.03% of samples contained no sequences (Tables S2A–S2E and Figures S4, S5 in SI1). In embryonic E13-nat samples, we revealed higher ASV diversity than in E0-eggs, with a total of 49 ASVs and mean 4.39 ASV per sample (Table S2D, S2E in SI1). Of these, 28 ASVs were unique to E13-nat, but these were covering only 2.3% of all reads (Figure S9A and see further panels B–E for genus-level comparisons and also for P2, SI1). The E13-nat samples were dominated by genera *Spingomonas*, *Mycobacterium*, *Methylotener*, and *Pseudomonas* (found in > 10 samples) and furthermore by *Cutibacterium*, *Stenotrophomonas*, *Enhydrobacter*, *Legionella*, *Blastocatellia*, *Brevundimonas*, and *Tepidionomonas* (found in ≥ 2 samples; Table 1; Fig. 4; Figures S4–S7 in SI1). *Clostridium* was revealed in one E13-nat sample. In addition, we were again able to detect the two putative pathogens *Dietzia* (two samples) and *Corynebacterium* (one sample) using the P2 approach. Of the total 49 ASVs detected in the E13-nat

samples, nine ASVs were shared with the E0 eggs, including the most common genera, such as *Cutibacterium*, *Mycobacterium*, and *Spingomonas* (Table 1; Figure S9A and also B–E for further comparisons, SI1).

The embryonic GIT samples derived from E0 eggs experimentally injected with *E. faecium* (E13-Ent) were dominated by the genus *Enterococcus*, as predicted, regardless of the dose applied (Figures S4, S7 in SI1). This demonstrates that our 16S rRNA gene metabarcoding approach is sensitive to the detection of bacteria even when they are initially present in low quantities in the eggs (≥ 1000 copies, the low dose of *E. faecium*). *In ovo* treatment also did not affect the microbiome of E13, as the control E13-PBS samples contained a similar composition of ASVs as E13-nat samples (Fig. 4A and B; Table S3 in SI2). Despite all our efforts to keep our procedure contamination-free, we observed a high abundance of *Enterococcus* in a few E13-PBS samples, possibly due to cross-contamination.

Female faecal samples generally had a much higher number of reads per sample than the E0-egg and E13-nat samples (Tables S2A–S2E in SI1). In total, 30 female faecal samples contained 49 ASVs, of which 34 ASVs were unique to faeces (this covered 28.7% of all sequences, Figures S9A in SI1). We observed high interindividual variability between females (Figures S5, S8A, S8B in SI1) and taxonomically different composition compared to all other biological sample types (Fig. 4). Among the most common taxa, *Mycoplasma*, *Clostridium*, *Escherichia*, *Lactococcus*, *Diplorickettsiaceae*, or *Ureoplasma* were detected. There were only six common ASVs shared between E0-egg and female samples (e.g. in genera *Clostridium*, *Devosia gracialis*, *Lactococcus*, *Mycobacterium*, and *Spingomonas*) and nine ASVs shared between E13-nat and female samples (e.g. in genera *Mycobacterium*, *Rhodococcus*, and *Spingomonas*; Tables S3A, S3B in SI2 and Figures S9A in SI1).

Similarity of microbial communities within- and between nests

Female faecal microbiota was not more similar to the microbiota of their own eggs (E0-egg) compared to foreign eggs (for Jaccard $W = 15\ 174$, $P = .726$ and Bray–Curtis $W = 15\ 152$, $P = .741$). On the other hand, we detected partially increased similarity between female faecal microbiota and microbiota of embryos in their nests (E13-nat) than would be expected by chance but only nonsignificantly for Jaccard ($W = 39\ 346$, $P = .086$) and not for Bray–Curtis ($W = 36\ 823$, $P = .368$). E13-nat within the same nest were no more similar in microbial composition than E13-nat between different nests (for Jaccard $W = 21\ 862$, $P = .021$ and Bray–Curtis $W = 35\ 762$, $P = .125$).

qPCR detection of potentially pathogenic bacteria

Unlike 16S rRNA gene metabarcoding, none of the 52 E0 egg samples were found to contain *Corynebacterium* and *Clostridium* by specific qPCR assays (see Table S8 in SI1 for comparison of NGS and qPCR results). In contrast, *Dietzia* was revealed by qPCR in four E0-egg samples (7.62%). However, *Dietzia* was also detected in INC and its frequency in E0-eggs was not statistically different from any type of negative control in E0-egg dataset (sample type: $P = .318$, Model 1, M1 in SI3, where see for full model details, and Figures S10A in SI1).

We also detected no *Corynebacterium* and *Clostridium* in any of the 66 E13-nat samples assessed. *Dietzia* was found by qPCR in only two E13-nat samples (i.e. 2.66%). Overall, *Dietzia* positivity was not statistically different from the negative controls (especially

Table 1. List of bacterial genera found in egg content (E0-egg) and embryonic intestine samples (E13-nat) in the great tit. The lowest taxonomic assignments to genera are shown where possible, based on the RDP classifier (Wang et al. 2007) and Silva reference database (Quast et al. 2013). For a complete list of all amplicon sequence variants (ASVs) and genera in all sample types, please see Table S3A and B, Supplementary Information S12). E0-egg – egg content samples at embryonic day 0 (N = 47), E13-nat – embryonic intestine sample at embryonic day 13 (N = 66), INC – negative controls of isolation (N = 26), NTC – negative control of PCR (N = 12). ID indicates the unique code of the given ASV. Presence of certain taxa for protocol 1 (P1; see methods for more details) is indicated by the number of positive samples. Decontam – genera revealed in the statistical analysis by Decontam are indicated (see Table S7, Supplementary Information S12 for a complete list of all contaminating ASVs revealed by Decontam analysis). Previously reported genera as common contaminants are cited: [1] Salter et al. (2014), [2] Eisenhofer et al. (2019), [3] Stinson et al. (2019). Genera identified only in E0-egg and E13-nat samples and not in negative controls are highlighted in grey.

ID	Genus	E0-egg	E13-nat	INC	NTC	Decontam	Published contaminants
ASV_10635	1174-901-12	0	0	1	0		
ASV_11627	<i>Arachidicoccus</i>	0	1	0	0		
ASV_10629	Beijerinckia	0	1	1	0		
ASV_10711	<i>Bradyrhizobium</i>	0	2	2	0		[1], [2], [3]
ASV_10591	<i>Brevundimonas</i>	0	3	0	0		[1], [2]
ASV_10053	Burkholderia-Caballeronia-Paraburkholderia	0	1	0	0		[1], [2]
ASV_11823	<i>Clostridium_sensu_stricto_1</i>	1	1	0	0		[2]
ASV_10865	<i>Craurococcus-Caldovatus</i>	0	1	0	0		[1]
ASV_11235	<i>Cutibacterium</i>	7	5	3	1	Yes	[3]
ASV_10844	<i>Devosia</i>	1	0	0	0		[1], [2]
ASV_9381	<i>Enhydrobacter</i>	1	2	1	0		[1], [2]
ASV_12311	<i>Enterococcus</i>	0	9	0	0		[2]
ASV_9291	<i>Haliangium</i>	0	1	0	0		
ASV_9636	KCM-B-112	0	0	1	0		
ASV_12263	<i>Lactococcus</i>	1	0	0	0		
ASV_9652	<i>Legionella</i>	0	2	0	0		
ASV_10652	Methylobacterium-Methylorubrum	1	0	1	0	Yes	[1], [2], [3]
ASV_10100	<i>Methylotenera</i>	0	20	3	0	Yes	
ASV_11228	<i>Microlunatus</i>	0	0	1	0		
ASV_11070	<i>Mycobacterium</i>	5	19	8	1	Yes	
ASV_11257	<i>Nocardioides</i>	0	1	0	0		
ASV_10567	<i>Novosphingobium</i>	0	1	0	0		[1], [2], [3]
ASV_9747	<i>Pseudomonas</i>	1	15	4	0		[1], [2]
ASV_9388	<i>Psychrobacter</i>	0	1	0	0		[1]
ASV_9841	<i>Ralstonia</i>	47	66	20	0		[1], [2], [3]
ASV_10991	<i>Rhodococcus</i>	0	5	1	0	Yes	[1], [3]
ASV_10372	Sphingomonadaceae	1	1	0	0		
ASV_10397	<i>Sphingomonas</i>	2	24	4	0	Yes	[1], [2], [3]
ASV_12175	<i>Staphylococcus</i>	1	1	0	0	Yes	[2]
ASV_10329	<i>Stenotrophomonas</i>	0	4	0	0		[1], [2]
ASV_12264	<i>Streptococcus</i>	0	0	1	0	Yes	[1], [2]
ASV_10170	<i>Tepidimonas</i>	0	2	0	0		
ASV_10678	Xanthobacteraceae	7	0	2	0		

from INC) in the E13-nat dataset (Sample type: $P = .154$, M2, SI3; Figures S10B and S11).

In female faeces, all *Corynebacterium* (25.00% of samples), *Clostridium* (90.63%), and *Dietzia* (90.63%) have been detected. However, in further statistical tests, only *Clostridium* and *Corynebacterium* showed significantly higher positivity in the faecal samples compared to the negative controls (sample type: $P = .004$, M3 and $P = .042$, M4, SI3; Figures S10C and S10D in S11, respectively) but not *Dietzia* (for sample type: $P = .103$, M5, SI3; Figures S10E and S11). Unlike in low-bacterial biomass samples (i.e. E0-egg and E13-nat), virtually all *Clostridium*-positive samples (i.e. 91.66%) revealed by NGS in females were also confirmed by the more sensitive qPCR, indicating that for the higher bacterial biomass samples both methods gave consistent results (Table S8 in S11).

Discussion

Using a combination of two 16S rRNA-based metabarcoding protocols and specific qPCR assays, in this study we described microbiome of the egg content, embryonic gut and female (maternal) faeces in a free-living passerine, the great tit. Importantly, unlike previous avian studies, we performed our PCR in technical replicates and prepared the sequencing runs for low- and high-bacterial biomass samples independently to minimize the risks of amplification bias and potential cross-contamination. Contrary to some previous studies in birds (e.g. Ding et al. 2017, Treveline et al. 2018), our results revealed negligible and inconsistent microbiota communities in eggs, with high proportions of potential contaminants. In embryonic samples, there were more bacterial ASVs, yet still these frequently represented putative contaminants. Of the three potentially pathogenic ASVs detected by our

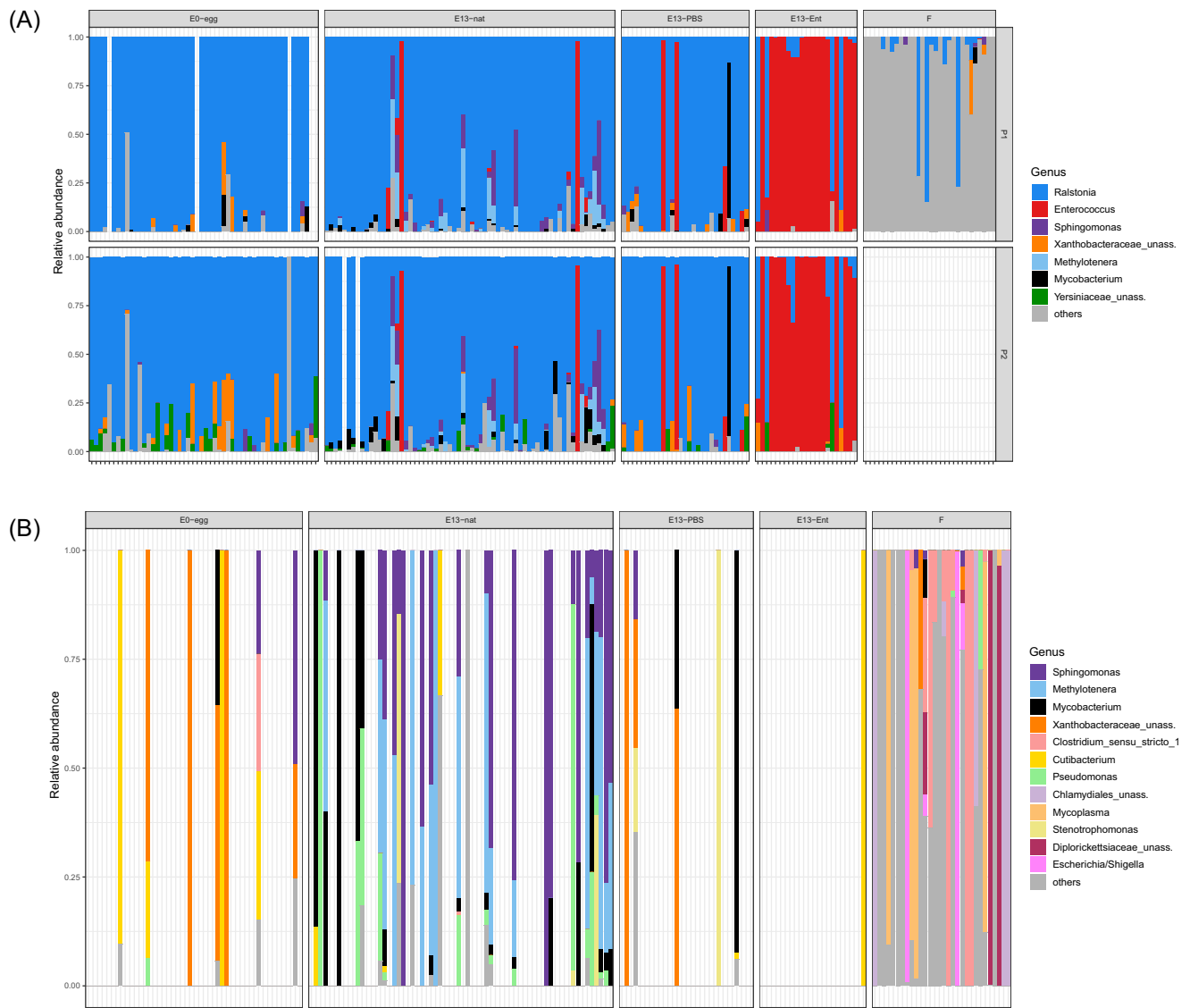


Figure 4. Relative abundances of bacterial taxa in egg and embryonic samples of the great tit (A) before excluding potentially contaminating *Ralstonia* and cross-contaminating *Enterococcus* for both protocols (P1 and P2) and (B) and after their exclusion for P1 only. Only the most abundant genera (relative abundance > 1%) are shown. Less abundant genera are included in the category ‘Other’. Results are shown for different amplification protocols (P1 and P2, see the section ‘Methods’ for more details). E0-egg—egg content sample at embryonic day 0, E13-nat—E13 intestinal sample from nonmanipulated egg, E13-Ent—E13 intestinal sample from an *Enterococcus*-treated egg, E13-PBS—E13 intestinal sample from a control PBS-injected egg, F—female faecal samples (amplified by P1 only). For the relative abundances in P2 after *Ralstonia* and *Enterococcus* exclusion, see [Figures S1 and S5 \(S1\)](#). High-resolution images of [Fig. 4\(A\)](#) and [\(B\)](#) with individual sample IDs can be found in [Supplementary Information 5\(A\)](#) and [\(B\)](#), respectively.

NGS metabarcoding in the E0-egg and E13-nat samples (*Corynebacterium*, *Clostridium*, and *Dietzia*), only *Dietzia* was confirmed by qPCR in E0-egg and E13-nat, which might support the putative vertical transfer.

E0-egg samples had only minimal traces of microbiota, as evidenced by the high proportion of samples without any sequences after the filtering and low mean ASV and number of reads per sample despite the increased number of preamplification PCR cycles. The E0-egg microbial communities were dominated by Xanthobacteraceae, *Cutibacterium*, *Mycobacterium*, and *Sphingomonas*. However, virtually all ASVs identified in E0-eggs were also detected in our negative controls (INCs and/ or NTCs) and most of them are also known as frequent contaminants of laboratory materials (see discussion below and [Table 1](#) for the detailed list of contaminants). Furthermore, our results indicate that the E0 mi-

crobiota in eggs from the same nest was not more similar than that of eggs from different nests. The qPCR validation of the NGS results showed that unlike *Dietzia*, neither *Corynebacterium* nor *Clostridium* were detected in any E0-egg samples. Therefore, our results suggest that the microbiota in great tit eggs at E0 is negligible. This is supported by (i) the lack of clearly distinguishable microbiota between the E0-egg and NTC and INC samples, (ii) inconsistency of technical duplicates, (iii) low ASV abundance and diversity, and (iv) high proportion of potential contaminants. The only exception supporting the possibility of vertical transfer in our data represents *Dietzia*, which was consistently identified in female faeces and in egg and embryonic samples, although it occurred also in the negative controls. Hence, while we cannot completely rule out some bacterial presence in the freshly laid eggs, our E0 eggs were nearly bacteria-free, which is consistent with the

sterile egg hypothesis. In contrast to our results, in E0 egg whites of the Korean chickens, diversified bacterial communities were reported, being dominated by *Pseudomonas* (65% of all reads), *Janthinobacterium*, Burkholderiales, *Flavobacterium*, *Stenotrophomonas*, *Acinetobacter*, Enterobacteriaceae, Comamonadaceae, and Xanthomonadaceae (Lee et al. 2019). Microbial communities in egg whites distinct from the negative controls were also described in four passerine species, yet in a very small dataset of only 11 eggs in total (Trevelline et al. 2018). Different microbial communities at E0 stage eggs between these studies might either indicate high interspecific variability or reflect insufficient control over the stochasticity in sequence data when replicates and/or negative controls are lacking.

The risk of potential contamination is a critical issue in low-bacterial biomass samples (Salter et al. 2014, Eisenhofer et al. 2019). The sample contamination may originate from different routes, including the environment in which samples are collected, DNA extraction kits, PCR chemicals and plastics, wet lab procedures, or cross-contamination between different samples (Eisenhofer et al. 2019). Contrary to mammalian studies where an ongoing debate whether human placenta and foetus are sterile (Perez-Muñoz et al. 2017) or not (Aagaard et al. 2014) has led to the application of very careful protocols assessing contamination risks (de Goffau et al. 2019), these caveats have rarely been addressed in birds. For instance, previous studies in chicken egg and embryonic microbiota did not include negative controls (Ding et al. 2017, 2022) or these were of insufficient type and number (Akinyemi et al. 2020), limiting inferences about microbiota profiles (but see Grond et al. 2017 and Lee et al. 2019). Unsurprisingly, the most abundant taxon in the chicken embryonic samples, *Halomonas* (Ding et al. 2017), was later indicated as a saline buffer contaminant of the DNA isolation kit (Lee et al. 2019). Our study is the first describing bacterial recruitment in bird embryos by adopting multiple protocols or the technical duplicate-based approach. Incorporating these measures and various bioinformatic filtering steps helped us to detect PCR amplification bias, which may be important in low-bacterial biomass and to remove many inconsistent ASVs that likely originated from environmental contaminants or as cross-contaminants. While our initial dataset consisted of 1382 ASVs, it was reduced to 128 ASVs after removing samples with inconsistent technical duplicates and identified contaminants. Studies without such measures identified up to an order of magnitude higher ASV diversity in chickens (Akinyemi et al. 2020, Ding et al. 2022). Despite that, we still did find some potentially contaminating ASVs identified based on the literature survey. Contaminating ASVs reported in previous microbial studies have been mostly soil- or water-dwelling bacteria and bacteria commonly associated with nitrogen fixation (Salter et al. 2014). Of these known contaminants (Salter et al. 2014, Eisenhofer et al. 2019, Stinson et al. 2019), we identified e.g. *Pseudomonas*, *Ralstonia*, *Rhodococcus*, *Sphingomonas*, and *Stenotrophomonas* in our E0-egg and E13-nat samples. In particular, *Ralstonia*, a common contaminant of DNA isolation kits that can pass even through a 20-nm bacterial filter (Sundaram et al. 1999), was highly dominating in our low-bacterial biomass samples, comprising 30.1% of our reads. Other contaminants, such as *Propionibacterium* or *Streptococcus* are common human skin-associated organisms (Byrd et al. 2018). Although the vast majority of ASVs observed in our E0-egg and E13-nat samples are putative contaminants, we cannot exclude the possibility that they also play some biological role in eggs and developing embryos, as some of them have been revealed in GIT of adult birds (e.g. *Mycobacterium* and *Pseudomonas*; Kropáčková et al. 2017a, 2017b). Therefore, we only conclude that their occurrence

in avian eggs is highly speculative. In addition, some of the identified genera are also known to be opportunistic pathogens in humans, such as some members of the genera *Mycobacterium* (Primm et al. 2004) or *Pseudomonas* (Kerr and Snelling 2009). Unfortunately, our 16S rRNA gene metabarcoding approach could not distinguish most ASVs at a taxonomic level lower than genus, while the potential pathogenicity of some ASVs could also be specific to species or strain (Pan et al. 2014).

Compared to E0-egg, the embryonic gut (E13-nat) microbiome was slightly more consistent, with higher ASV diversity and the number of reads per sample, though 53.03% of samples contained no sequences after the data quality filtering. Of the 49 ASVs detected in the E13-nat samples, nine ASVs were shared with the E0 eggs, including the most common genera, such as *Cutibacterium*, *Mycobacterium*, and *Sphingomonas*. Yet, again, all genera in E13-nat except for *Stenotrophomonas*, *Legionella*, *Blastocatellia*, and *Tepidionomonas* were also detected in our negative controls. The lack of higher microbial similarity between embryos from the same nest indicates rather stochastic recruitment of microbiota at E13. Of the genera and higher taxa found in our great tit E13-nat samples, nine also inhabited the chicken embryos: *Pseudomonas*, *Sphingomonas*, *Bradyrhizobium*, *Rhodococcus*, *Staphylococcus*, *Pseudomonas*, Burkholderiales, *Stenotrophomonas*, and Enterobacteriaceae (Ding et al. 2017, Lee et al. 2019). Despite the sharing of these few genera between chicken and great tit embryos, overall abundances and taxonomical diversities were much lower in the great tits (Ding et al. 2017, 2022; Lee et al. 2019). On the one hand, this could reflect interspecific differences, as the embryos of the altricial great tits are less developed than those of the precocial chickens, on the other hand, the differences in the results may also be due to differences in the NGS protocols used (see above). Our results, therefore, better correspond with those of Grond et al. (2017) who combined NGS and qPCR protocols to detect only negligible microbiota in the embryonic GIT of two Arctic superprecocial shorebirds, the dunlin (*Calidris alpina*) and the sandpiper (*Calidris pusilla*), shortly before hatching, with stochastic communities skewed towards dominance of *Clostridia* and *Gammaproteobacteria*.

Using metabarcoding we detected potentially pathogenic genera *Corynebacterium*, *Clostridium*, and *Dietzia* both in E0-egg and E13-nat, and partially also in female faecal samples, suggesting their putative vertical transfer. Unlike the sequencing, no *Corynebacterium* and *Clostridium* were detected by qPCR both in E0-egg and E13-nat. Importantly, *Dietzia* was revealed by qPCR in 7.69% of E0-eggs and in 3.03% of E13-nat samples. This suggests that *Dietzia* might be vertically transmitted at a very low frequency. Similarly, the putative vertical transmission of pathogenic bacteria from infected mothers to their eggs was suggested in both domestic (e.g. for *Salmonella* in chickens; Keller et al. 1995, Gantois et al. 2009, Pedroso 2009) and wild birds (e.g. for *Nisseria* in the greater white-fronted goose; Hansen et al. 2015). As we found almost no microbiota in E0-egg samples but a somewhat more diversified microbiota in E13-nat (with limited overlap with E0-egg samples), we think that bacterial trans-shell migration in E13-nat seems to have a stronger effect than vertical transfer, yet it is still rare in the great tit. As described previously (Kropáčková et al. 2017a), we found rich bacterial communities in the female faecal microbiome. However, the fact that these were taxonomically very different from those in the E0-egg and E13-nat samples might suggest that the original embryonic ASVs are either replaced by other ASVs after hatching or exist in the adults only in very low, undetectable amounts. Our results thus suggest that GM is predominantly formed after hatching in passerines. This is also supported by the high ontogenetic dynamics of passerine GM during

the nestling period (e.g. Kreisinger et al. 2017, Teysier et al. 2018, Chen et al. 2020).

Conclusion

In our study combining 16S rRNA gene metabarcoding with targeted qPCR, we have revealed only nearly sterile eggs in the great tit. We found stronger support for the role of *bacterial trans-shell migration* than *vertical transfer* during embryogenesis, forming simple and low-abundant bacterial communities in some (but not all) tit embryos. However, all effects were weak, suggesting that GIT microbiota in passerines mostly forms only after hatching. Further careful investigation across avian phylogeny is needed to determine whether the differences in gut microbiota recruitment between passerines and chickens are due to species-specific life-history traits or methodological caveats between studies. Our study highlights the importance of using technical PCR duplicates and internal controls to eliminate stochastic noise and contamination in sequencing data, which are particularly common in studies with low-microbial biomass. Our results indicate that further studies in species with potentially more abundant microbial communities in embryonic intestine (e.g. chickens) should determine whether bacteria in bird eggs are viable (e.g. by labelling dead/live bacteria) and metabolically active (e.g. by metatranscriptomics and RT-qPCR) and assess their impact on immune gene expression. This can also be done using microbial culturomics, which combines high-throughput culturing with species identification by MALDI-TOF (Lagier et al. 2018). Elucidating the mechanisms of GIT microbiota establishment in early ontogeny can improve our understanding of parental effects in birds and contribute both to basic knowledge of host-microbe evolutionary ecology as well as zoohygienic and veterinary measures that minimize the risks of disease transmission.

Data accessibility

All supporting data, including the original metadata, are presented in the [Supporting Information \(SI1–5; see the list below\)](#).

SI1 – Supplementary methods, figures and tables (main supplement)

SI2 – Metabarcoding and qPCR sample metadata, metabarcoding abundance table for amplicon sequence variants (ASVs) and genera

SI3 – Full Generalized Linear Models (GLMs) details

SI4 – Accession numbers for the individual samples stored at European Nucleotide Archive

SI5A and B – High-resolution images of Figures 4A and B with individual sample IDs

Sequencing data are available in the European Nucleotide Archive under the project accession number PRJEB61256. The accession numbers for the individual samples can be found in SI4.

Author contributions

Martin Těšický (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing), Lucie Schmiedová (Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing), Tereza Krajzingrová (Funding acquisition, Investigation, Writing – review & editing), Mercedes Gomez Samblas (Methodology, Supervision, Writing – review & editing), Petra Bauerová (Investigation, Writing – review & editing),

Jakub Kreisinger (Conceptualization, Formal analysis, Methodology, Software, Supervision, Visualization, Writing – review & editing), and Michal Vinkler (Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing).

Ethical approval

The research was carried out under the applicable laws of the Czech Republic and the European Union. The experiment was approved by the Environmental Protection Department of Prague City Hall (permit number S-MHMP-1061728/2010/OPP-V-790/R-235/Bu) and the Ministry of the Environment of the Czech Republic (permit number 22003/ENV/16-1009/630/16). Permission for the capture and ringing of adult birds was granted by the Bird Ringing Centre of the National Museum in Prague.

Acknowledgement

We thank Dan Divín and Jiří Eliáš for their help with field sampling, Dagmar Čížková, Anna Bryjová, and Zuzana Świederská for their help and advice with molecular-genetics analysis. We are also grateful to Natalie Salem for her help with language correction.

Supplementary data

Supplementary data is available at [FEMSEC Journal](#) online.

Conflict of interest: The authors declare that they have no conflict of interest.

Funding

This study was supported by Charles University (grants numbers GAUK 1158217, UNCE 204069, and START/SCI/113 with reg. no. CZ.02.2.69/0.0/0.0/19_073/0016935) and Institutional Research Support (number 260684/2023). Computational resources were supplied by the project ‘e-Infrastruktura CZ’ (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures supported by the Ministry of Education, Youth and Sports of the Czech Republic (MEYS CR). We acknowledge the CF Genomics of CEITEC supported by the NCMG research infrastructure (LM2015091 funded by MEYS CR) for their support in obtaining the scientific data presented in this paper.

References

- Aagaard K, Ma J, Antony KM et al. The placenta harbors a unique microbiome. *Sci Transl Med* May 2014;**21**:237–65.
- Akinyemi FT, Ding J, Zhou H et al. Dynamic distribution of gut microbiota during embryonic development in chicken. *Poult Sci* 2020;**99**:5079–90.
- Bäckhed F, Ley RE, Sonnenburg JL et al. Host-bacterial mutualism in the human intestine. *Science* 2005;**307**:1915–20.
- Barbosa A, Palacios MJ. Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Pol Biol* 2009;**32**:1095–115.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995;**57**:289–300.

- Benskin CMWH, Wilson K, Jones K et al. Bacterial pathogens in wild birds: a review of the frequency and effects of infection. *Biol Rev* 2009;**84**:349–73.
- Bi Y, Tu Y, Zhang N et al. Multiomics analysis reveals the presence of a microbiome in the gut of fetal lambs. *Gut* 2021;**70**:853–64.
- Bruce J, Drysdale EM. Trans-shell transmission BT—microbiology of the avian egg. In: Board RG, Fuller R (eds.), *Microbiology of the Avian Egg*. Boston: Springer, 1994, 63–91.
- Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. *Nat Rev Micro* 2018;**16**:143–55.
- Callahan BJ, McMurdie PJ, Rosen MJ et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;**13**:581–3.
- Campos-Cerda F, Bohannon BJM. The nidobiome: a framework for understanding microbiome assembly in neonates. *Trends Ecol Evol* 2020;**35**:573–82.
- Canty A, Ripley BD. boot: bootstrap R (S-Plus) functions. CRAN, 2021.
- Cuperus T, Coorens M, van Dijk A et al. Avian host defense peptides. *Dev Compar Immunol* 2013;**41**:352–69.
- D’Alba L, Shawkey MD. Mechanisms of antimicrobial defense in avian eggs. *J Ornithol* 2015;**156**:399–408.
- Davis NM, DiM P, Holmes SP et al. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018;**6**:1–14.
- de Goffau MC, Lager S, Sovio U et al. Human placenta has no microbiome but can contain potential pathogens. *Nature* 2019;**572**:329–34.
- Ding J, Dai R, Yang L et al. Inheritance and establishment of gut microbiota in chickens. *Front Microbiol* 2017;**8**:1–11.
- Ding P, Liu H, Tong Y et al. Developmental change of yolk microbiota and its role on early colonization of intestinal microbiota in chicken embryo. *Animals* 2022;**12**:1–15.
- Dorn-In S, Bassitta R, Schwaiger K et al. Specific amplification of bacterial DNA by optimized so-called universal bacterial primers in samples rich of plant DNA. *J Microbiol Methods* 2015;**113**:50–6.
- Edgar RC, Haas BJ, Clemente JC et al. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;**27**:2194–200.
- Eisenhofer R, Minich JJ, Marotz C et al. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol* 2019;**27**:105–17.
- Funkhouser LJ, Bordenstein SR Mom knows best: the universality of maternal microbial transmission. *PLoS Biol* 2013;**11**:1–9.
- Gantois I, Ducatelle R, Pasmans F et al. Mechanisms of egg contamination by *Salmonella enteritidis*: review article. *FEMS Microbiol Rev* 2009;**33**:718–38.
- Grond K, Lanctot RB, Jumpponen A et al. Recruitment and establishment of the gut microbiome in arctic shorebirds. *FEMS Microbiol Ecol* 2017;**93**:1–9.
- Hansen CM, Meixell BW, Van Hemert C et al. Microbial infections are associated with embryo mortality in Arctic-nesting geese. *Appl Environ Microb* 2015;**81**:5583–92.
- Honda K, Littman DR. The microbiome in infectious disease and inflammation. *Annu Rev Immunol* 2012;**30**:759–95.
- Chen CY, Chen CK, Chen YY et al. Maternal gut microbes shape the early-life assembly of gut microbiota in passerine chicks via nests. *Microbiome* 2020;**8**:1–13.
- Jiang H, Lei R, Ding SW et al. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf* 2014;**15**:1–12.
- Keller LH, Benson CE, Krotec K et al. *Salmonella enteritidis* colonization of the reproductive tract and forming and freshly laid eggs of chickens. *Infect Immun* 1995;**63**:2443–9.
- Kennedy KM, de Goffau MC, Perez-Muñoz ME et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature* 2023;**613**:639–49.
- Kerr KG, Snelling AM. *Pseudomonas aeruginosa*: a formidable and ever-present adversary. *J Hosp Infect* 2009;**73**:338–44.
- Kizerwetter-Świda M, Binek M. Bacterial microflora of the chicken embryos and newly hatched chicken. *J Anim Feed Sci* 2008;**17**:224–32.
- Klindworth A, Pruesse E, Schweer T et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013;**41**:e1–.
- Koerner RJ, Goodfellow M, Jones AL. The genus *Dietzia*: a new home for some known and emerging opportunist pathogens. *FEMS Immunol Med Microbiol* 2009;**55**:296–305.
- Kreisinger J, Kropáčková L, Petrželková A et al. Temporal stability and the effect of transgenerational transfer on fecal microbiota structure in a long distance migratory bird. *Front Microbiol* 2017;**8**:1–19.
- Kropáčková L, Pechmanová H, Vinkler M et al. Variation between the oral and faecal microbiota in a free-living passerine bird, the great tit (*Parus major*). *PLoS ONE* 2017a;**12**:e0179945. <https://doi.org/10.1371/journal.pone.0179945>.
- Kropáčková L, Těšický M, Albrecht T et al. Codiversification of gastrointestinal microbiota and phylogeny in passerines is not explained by ecological divergence. *Mol Ecol* 2017b;**26**:5292–304.
- Lagier JC, Dubourg G, Million M et al. Culturing the human microbiota and culturomics. *Nat Rev Micro* 2018;**16**:540–50.
- Lee SW, La TM, Lee HJ et al. Characterization of microbial communities in the chicken oviduct and the origin of chicken embryo gut microbiota. *Sci Rep* 2019;**9**:1–11.
- Liong JWW, Frank JF, Bailey S. Visualization of eggshell membranes and their interaction with *Salmonella enteritidis* using confocal scanning laser microscopy. *J Food Prot* 1997;**60**:1022–8.
- Liu C, Cui Y, Li X et al. Microeco: an R package for data mining in microbial community ecology. *FEMS Microbiol Ecol* 2021;**97**:1–9.
- Lüdecke D. ggeffects: create tidy data frames of marginal effects for “ggplot” from model outputs. GitHub, 2018.
- Lunam CA, Ruiz J. Ultrastructural analysis of the eggshell: contribution of the individual calcified layers and the cuticle to hatchability and egg viability in broiler breeders. *Br Poult Sci* 2000;**41**:584–92.
- Mann K. The chicken egg white proteome. *Proteomics* 2007;**7**:3558–68.
- Mishra A, Lai GC, Yao LJ et al. Microbial exposure during early human development primes fetal immune cells. *Cell* 2021;**184**:3394–409.e20.
- Olowo-okere A, Ibrahim YKE, Lo CI et al. *Bhargavaea massiliensis* sp. nov. and *Dietzia massiliensis* sp. nov., novel bacteria species isolated from human urine samples in Nigeria. *Curr Microbiol* 2022;**79**:1–8.
- Ost KS, Round JL. Communication between the microbiota and mammalian immunity. *Annu Rev Microbiol* 2018;**72**:399–422.
- Pan X, Yang Y, Zhang JR. Molecular basis of host specificity in human pathogenic bacteria. *Emerg Microbes Infect* 2014;**3**:1.
- Pedroso AA. Which came first: the egg or its microbiota?. *Poult Inf Prof* 2009;**48**:1–5.
- Perez-Muñoz ME, Arrieta MC, Ramer-Tait AE et al. A critical assessment of the “sterile womb” and “in utero colonization” hypotheses: implications for research on the pioneer infant microbiome. *Microbiome* 2017;**5**:1–19.
- Primm TP, Lucero CA, Falkinham JO. Health impacts of environmental *Mycobacteria*. *Clin Microbiol Rev* 2004;**17**:98–106.
- Quast C, Pruesse E, Yilmaz P et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590–6.

- R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2017. <https://www.R-project.org/>.
- Rackaityte E, Halkias J, Fukui EM et al. Viable bacterial colonization is highly limited in the human intestine in utero. *Nat Med* 2020;**26**:599–607.
- Risely A, Waite DW, Ujvari B et al. Active migration is associated with specific and consistent changes to gut microbiota in *Calidris* shorebirds. *J Anim Ecol* 2018;**87**:428–37.
- Roto SM, Kwon YM, Ricke SC. Applications of in ovo technique for the optimal development of the gastrointestinal tract and the potential influence on the establishment of its microbiome in poultry. *Front Vet Sci* 2016;**3**:1–13.
- Ryan MP, Adley CC. *Ralstonia* spp.: emerging global opportunistic pathogens. *Eur J Clin Microbiol Infect Dis* 2014;**33**:291–304.
- Salter SJ, Cox MJ, Turek EM et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;**12**:1–12.
- Stinson LF, Keelan JA, Payne MS. Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome analyses. *Lett Appl Microbiol* 2019;**68**:2–8.
- Strandwitz P. Neurotransmitter modulation by the gut microbiota. *Brain Res* 2018;**1693**:128–33.
- Sundaram S, Auriemma M, Howard G et al. Application of membrane filtration for removal of diminutive bioburden organisms in pharmaceutical products and processes. *PDA J Pharm Sci Technol* 1999;**53**:186–201.
- Svensson L, Baker K. *Identification Guide to European Passerines*. 4th edn. Stockholm: British Trust for Ornithology, 1992.
- Těšický M, Krajzingrová T, Eliáš J et al. Inter-annual repeatability and age-dependent changes in plasma testosterone levels in a longitudinally monitored free-living passerine bird. *Oecologia* 2022;**198**:53–66.
- Těšický M, Krajzingrová T, Šwiderská Z et al. Longitudinal evidence for immunosenescence and inflammaging in free-living great tits. *Exp Gerontol* 2021;**154**:111527.
- Teyssier A, Lens L, Matthysen E et al. Dynamics of gut microbiota diversity during the early development of an avian host: evidence from a cross-foster experiment. *Front Microbiol* 2018;**9**. <https://doi.org/10.3389/fmicb.2018.01524>.
- Trevelline BK, MacLeod KJ, Knutie SA et al. In ovo microbial communities: a potential mechanism for the initial acquisition of gut microbiota among oviparous birds and lizards. *Biol Lett* 2018;**14**:3–6.
- Tsiodras S, Kelesidis T, Kelesidis I et al. Human infections associated with wild birds. *J Infect* 2008;**56**:83–98.
- Van Veelen HPJ, Salles JF, Tieleman BI. Microbiome assembly of avian eggshells and their potential as transgenerational carriers of maternal microbiota. *ISME J* 2018;**12**:1375–88.
- Walker RW, Clemente JC, Peter I et al. The prenatal gut microbiome: are we colonized with bacteria in utero?. *Pediatr Obesity* 2017;**12**:3–17.
- Wang Q, Garrity GM, Tiedje JM et al. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microb* 2007;**73**:5261–7.
- Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.
- Wright ES. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinf* 2015;**16**:1–14.