Multiple host transfers, but only one successful lineage in a continent-spanning emergent pathogen

Wesley M. Hochachka1, André A. Dhondt1, Andrew Dobson2, Dana M. Hawley3, David H. Ley4 and Irby J. Lovette1

1Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA
2Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
3Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA
4Department of Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27607, USA

Emergence of a new disease in a novel host is thought to be a rare outcome following frequent pathogen transfers between host species. However, few opportunities exist to examine whether disease emergence stems from a single successful pathogen transfer, and whether this successful lineage represents only one of several pathogen transfers between hosts. We examined the successful host transfer and subsequent evolution of the bacterial pathogen *Mycoplasma gallisepticum*, an emergent pathogen of house finches (*Haemorhous* (formerly *Carpodacus*) *mexicanus*). Our principal goals were to assess whether host transfer has been a repeated event between the original poultry hosts and house finches, whether only a single host transfer was ultimately responsible for the emergence of *M. gallisepticum* in these finches, and whether the spread of the pathogen from east to west across North America has resulted in spatial structuring in the pathogen. Using a phylogeny of *M. gallisepticum* based on 107 isolates from domestic poultry, house finches and other songbirds, we infer that the bacterium has repeatedly jumped between these two groups of hosts but with only a single lineage of *M. gallisepticum* persisting and evolving in house finches; bacterial evolution has produced monophyletic eastern and western North American subclades.

1. Introduction
Transfer of a pathogen to a novel host is a common mechanism by which diseases emerge [1–4]. The first step in a host-transfer emergence of disease is the transmission of the pathogen between its original host and the novel host. Prior work suggests that following an initial jump to a new host a pathogen will often not be well-adapted, and that in many cases, transmission will be restricted to a few stuttering chains of infected individuals that ultimately lead to failure of the pathogen to establish in the novel host species [3,5,6]. Before transfer to a novel host can be followed by independent evolution, it is likely that repeated asymmetrical transfers of pathogens have happened in a sort of source–sink dynamics [5]; eventually, a ‘lucky mutant’ is transferred that can sustain itself in the new host. By contrast, anecdotal evidence exists for repeated introduction of some pathogens into novel hosts [7,8], and repeated transfers are implicit in our understanding of the major evolutionary shifts of human influenza [9].

Few large archives of samples exist for pathogens with non-human hosts [10–12], and to the best of our knowledge none have been used to describe the emergence of disease in a novel host. If archived specimens are available, then phylogenetic reconstructions can be used to make inferences about disease emergence and subsequent dynamics. Phylogenies make it possible to infer that disease emergence resulted from a single successful host jump, when only a single clade of pathogen is found in extensive sampling of the pathogens in a novel host. Archived specimens further allow the examination of subsequent
evolution of the pathogen in a novel host, with respect both to longer-term spatial and temporal alterations in genomes [10,11] and with finer-resolution changes in genetic diversity at individual geographical locations [13,14].

One non-human system in which pathogen genotypes can be examined throughout the process of emergence in a new host is the bacterium Mycoplasma gallisepticum and its novel host the house finch (Haemorhous formerly Carpodacus mexicanus). Mycoplasma gallisepticum is a worldwide pathogen of domestic poultry and is relatively taxon-specific in the hosts in which it commonly causes disease (Galliformes). The emergence of M. gallisepticum as a pathogen in house finches in 1993–1994 [15] was highly unusual and probably involved substantial differentiation between house finch and poultry lineages. Studies of a limited number of isolates suggest M. gallisepticum emergence in house finches entailed rapid differentiation of the bacterium during the colonization of house finches [16,17]; this is consistent with a single (point-source) origin. Nevertheless, anecdotal evidence [16,18]—one subsequent isolate each of apparent finch lineage and poultry lineage M. gallisepticum crossing to poultry and house finches, respectively—also suggested multiple transfers of M. gallisepticum between galliforms and finches. Following the emergence of M. gallisepticum in house finches, other wild songbirds (order Passeriformes) have also found to be infected with M. gallisepticum; these appear to have originated from house finches [19,20]. No large-scale comparisons have been made of M. gallisepticum lineages in poultry, house finches and other songbirds to validate the hypotheses of a point-origin and single successful transfer of the pathogen in house finches and subsequent transfer from house finches to other songbirds.

Multiple lines of evidence suggest that substantial spatial and temporal genetic variation exists in M. gallisepticum that has become established in house finches, and that this variation may underlie systematic phenotypic changes in virulence through space and time. The bacteria spread in house finches from the point of emergence in the mid-Atlantic United States (close to Baltimore) to the Pacific coast [21], apparently across the northern Great Plains of the United States [22,23].

Work with a small number of isolates showed that some genetic differentiation has occurred [16,17,24] and suggests that this genetic variation may be linked with systematic directional phenotypic changes in virulence. Experimental studies suggest that evolution of increasing virulence through time in lineages of M. gallisepticum present in eastern North America, as well as a drop in virulence with dispersal to western North America [25–27]. However, it is currently unclear whether eastern and western isolates of M. gallisepticum represent distinct phylogenetic clades as a result of a single pathogen dispersal from east to west. A clearer understanding of spatial and temporal evolution in virulence of M. gallisepticum will require the examination of the evolutionary relationships among a larger number of isolates of M. gallisepticum in house finches and other songbirds.

In this paper, we take advantage of a large collection of M. gallisepticum isolates from poultry, house finches and other songbirds in order to examine the phylogenetic relationships among bacterial isolates collected from these hosts. We ask the following questions: (i) is there evidence of repeated transfer of M. gallisepticum between domestic poultry and house finches? (ii) irrespective of repeated host transfers, is there a single dominant clade of M. gallisepticum in house finches and other songbirds that is distinct from the M. gallisepticum circulating in domestic poultry? (iii) did western North American isolates derive from eastern ones, consistent with a single host jump from poultry and later colonization of western North America? and (iv) when songbirds other than house finches are infected by M. gallisepticum, is the bacteria consistently from the same clade that infects house finches? While the details of our results are specific to this pathogen–host system, they provide important insights that are relevant to patterns of evolution that will occur in other pathogens, and more generally to invasive alien species [28] that have successfully colonized a novel niche.

2. Material and methods

The bacterial isolates included in our study were derived from one of three general sources: (i) field sampling of wild house finches and other songbirds collected during the study of M. gallisepticum dynamics, (ii) isolates from diseased house finches or other songbirds collected from wildlife rehabilitation facilities, and (iii) isolates from poultry on farms collected in support of disease control measures. Samples from all three sources were collected non-systematically both temporarily and spatially, and the sources vary in the precision with which collection location was recorded or the accuracy with which locations could be determined (detailed below). The non-systematic spatial and temporal sampling does not allow these data to be used for quantifying distributions of bacterial haplotypes or rates of evolution, but we have no expectation that this would bias assessments related to the questions that we have outlined in §1.

When wild house finches and other songbirds were trapped in the course of work (New York, Georgia, Wisconsin, Alabama) to identify the prevalence of M. gallisepticum infections (source 1), samples were collected following procedures as described in Hawley et al. [26], for subsequent culture and identification as described in the electronic supplementary material, section Methods. We also worked with wildlife rehabilitation centres in California, Virginia and North Carolina to obtain samples from house finches and other songbirds that were brought into these centres, presumably from nearby locations (source 2). Locations of sample collection are shown in figure 1a, and years of collection in figure 1b. Owing to the difficulty in successfully culturing field isolates, our collection protocol evolved over the course of study. Ultimately, we had our highest success when conjunctival swabs were placed into universal transport medium (Becton–Dickenson/Copan; Sparks, MD, USA) and shipped overnight on cold packs to NC State University College of Veterinary Medicine (NCSU CVM; Raleigh, NC, USA). Owing to improvements in our culture success with field-collected samples, we have a better representation of songbird isolates from the late 2000s and early 2010s than in prior years (figure 1b). Overall, the songbird isolates examined in this study represent approximately 52 per cent of those available to us, and were chosen to be representative of the dates, locations and hosts of isolates available. Six of eight of the M. gallisepticum house finch isolates used in our experimental studies of virulence [26,27] were examined in this study, and their identities are noted in the electronic supplementary material, table S1. Samples from all but one of the songbird isolates came from the archive maintained by D.H.L. at NC State University (see the electronic supplementary material, table S1). The remaining isolate, collected from a house finch in Ithaca, NY, was cultured in Ithaca immediately prior to sequencing.

The sample isolates of M. gallisepticum from poultry (source 3) were selected from a larger archive maintained by D.H.L. This archive of poultry isolates has been built in the course of D.H.L.’s diagnostic work in support of disease control on poultry...
Methods for laboratory culture and identification are described in Ley et al. [15]. *Mycoplasma gallisepticum* is a nutritionally fastidious and labile bacterium, making the isolation and creation of an archival library challenging. Each of the archived isolates was cultured in Frey’s mycoplasma media (broth and agar) with 15% swine serum medium [29], and colonies on agar identified as *M. gallisepticum* by direct immunofluorescence. Aliquots of low passage broth cultures were frozen at −70°C. See electronic supplementary material, section Methods for further information about culturing *M. gallisepticum*.

All molecular genetic work was carried out at Cornell University, with samples from NC State University shipped to Cornell on dry ice. We based our phylogeny of *M. gallisepticum* on partial sequencing of 107 isolates (60 from house finches and other songbirds, and 47 from domestic poultry). We selected regions of the genome to sequence without reference to functional genes (see below for details), creating data that are appropriate for examining patterns of genetic relatedness among isolates. This approach allows a broad survey of isolates that complements the whole-genome sequencing of a far smaller set of isolates that is needed to suggest candidate genes under selection during and subsequent to the establishment of *M. gallisepticum* in house finches and related species. Our research collaboration’s whole-genome sequencing is reported on elsewhere [24], and has identified variation in multiple genes associated with levels of virulence [24], one of these discussed in greater detail by Delaney et al. [17].

As the phylogenetic component of this survey was initiated before full DNA sequences were available for any *M. gallisepticum*, we initially used DNA from the index isolate of *M. gallisepticum* in house finches [15] to construct a small-insert genomic library from which we generated a panel of anonymous nuclear sequence markers via direct Sanger sequencing of the Mycoplasma genomic inserts. We designed and optimized 13 PCR primer pairs from within a corresponding set of *M. gallisepticum* insert sequences, each targeting a region of 559–869 nucleotides. See the electronic supplementary material, section Methods for more information. Comparisons against subsequently published genomes of *M. gallisepticum* confirmed that all of these regions aligned to, and were distributed throughout, the *M. gallisepticum* genome. Location in a reference *M. gallisepticum* genome of each amplified sequence, gene identity and related information are included in the electronic supplementary material, table S2; we have no reason to suspect that any sequenced regions are under selection in a way that would bias our phylogenetic results. All sequence data have been archived in GenBank (accession numbers KC995288–KC996678, inclusive); electronic supplementary material, table S3 lists the isolate and primers used to obtain each GenBank sequence, with isolate and primer names and further information provided in the electronic supplementary material, tables S1 and S2.

Genomic DNA was purified using a DNAeasy kit (Qiagen, Valencia, CA, USA), and all 13 regions were PCR amplified and sequenced from all 107 isolates using standard methods. Sequences were checked using Sequencer v. 4.6 (GeneCodes, Ann Arbor, MI, USA), and all sequences for each region were aligned readily by eye. The combined length of all aligned regions was 8399 nucleotides from a total genome slightly under one million base pairs in length [30].

The phylogenetic relationships among all of the isolates and associated posterior probability scores for branches within this phylogeny were reconstructed using Bayesian multi-chain MCMC analyses as implemented in MrBayes v. 3.1.2 [31,32]. Analyses were conducted on the concatenated sequences from all 13 regions, as these have a shared phylogenetic history in a clonal organism such as *M. gallisepticum*. MrBayes analyses were run for 25 million generations under a globally applied GTR + G + I substitution model. Trees were sampled every 1000 generations, and runs used two simultaneous Markov chains. Convergence was assessed using the average standard deviation of split frequencies;
obtained from house finches and other songbirds. (Figure 2 )

When a pathogen transfers to a novel host, the expectation is that the process involves: (i) repeated but typically short-lived introductions of a pathogen into a novel host until one introduction is either preceded or followed by adaptations that allow persistence, and then (ii) further adaptation of the pathogen to its new host [5,6]; these adaptations potentially restrict the pathogen’s ability to grow in its original host species. The phylogeny produced from our collection of isolates is consistent with both aspects of this general model for emergence of pathogens in novel hosts. Even with genotyping M. gallisepticum from a small proportion of all birds infected since this disease emerged in house finches, we found two haplotypes (‘Poultry_03.4’ and ‘Finch_NY01.1’ in figure 2) that imply the existence of continued exchanges of house finch clade of M. gallisepticum (figure 2b). We used TCS 1.21 [33] to construct this haplotype network.

3. Results and discussion

When a pathogen transfers to a novel host, the expectation is that the process involves: (i) repeated but typically short-lived introductions of a pathogen into a novel host until one introduction is either preceded or followed by adaptations that allow persistence, and then (ii) further adaptation of the pathogen to its new host [5,6]; these adaptations potentially restrict the pathogen’s ability to grow in its original host species. The phylogeny produced from our collection of isolates is consistent with both aspects of this general model for emergence of pathogens in novel hosts. Even with genotyping M. gallisepticum from a small proportion of all birds infected since this disease emerged in house finches, we found two haplotypes (‘Poultry_03.4’ and ‘Finch_NY01.1’ in figure 2) that imply the existence of continued exchanges of M. gallisepticum both from poultry to songbirds as well as from songbirds to poultry. Our phylogeny also shows a highly supported divergence between the haplotypes of

![Figure 2](http://rspb.royalsocietypublishing.org/)

Figure 2. Phylogenetic relationships among haplotypes of M. gallisepticum isolated from (a) all songbird and poultry haplotypes used in our study, and (b) isolates obtained from house finches and other songbirds. (a) All haplotypes recovered from house finches and other songbirds are indicated in red, and all haplotypes recovered from poultry are indicated in black. Note the presence of one ‘finch’ isolate recovered from domestic poultry and one ‘poultry’ isolate recovered from a house finch. The network diagram in (b) indicates a separation between haplotypes from California (in yellow) and those from eastern North America (open circles). The black-filled circle indicates the ‘house finch’ haplotype recovered from domestic turkeys in North Carolina, which clearly derived from the index isolate first recovered from a house finch in 1994 and from which a second distinct haplotype (found in a house finch) was in turn derived. Grey text in (b) denotes isolates from American goldfinch (‘AMGO’) or evening grosbeak (Coccothraustes vespertinus; ‘EVGR’), with only isolate ‘Finch_NC98.1’ being unique to a songbird other than house finches. Each dot on a connecting line and larger labelled circle represents a single nucleotide substitution. Circle areas monotonically increase with the number of isolates belonging to each haplotype. In both (a) and (b), all haplotypes from house finches and other songbirds are labelled starting with ‘finch’, followed by the two-letter code for the state in which the haplotype was first isolated, the last two digits of the calendar year in which the haplotype was recovered, and a single arbitrary number unique to that haplotype. All haplotypes from domestic poultry are labelled with ‘Poultry’, the last two digits of the calendar year of the first isolate identified with the haplotype and then an arbitrary number.

The results that (figure 2b) revealed that differences among the isolates from songbirds were minor in comparison with the variation among M. gallisepticum from poultry. Because of this similarity among songbird isolates, a more appropriate method for describing the associations among the isolates was to use statistical parsimony to create a haplotype network describing the relationships within the recently diverged finch clade of M. gallisepticum (figure 2b). We used TCS 1.21 [33] to construct this haplotype network.

The diagram in (b) indicates a separation between haplotypes from California (in yellow) and those from eastern North America (open circles). The black-filled circle indicates the ‘house finch’ haplotype recovered from domestic turkeys in North Carolina, which clearly derived from the index isolate first recovered from a house finch in 1994 and from which a second distinct haplotype (found in a house finch) was in turn derived. Grey text in (b) denotes isolates from American goldfinch (‘AMGO’) or evening grosbeak (Coccothraustes vespertinus; ‘EVGR’), with only isolate ‘Finch_NC98.1’ being unique to a songbird other than house finches. Each dot on a connecting line and larger labelled circle represents a single nucleotide substitution. Circle areas monotonically increase with the number of isolates belonging to each haplotype. In both (a) and (b), all haplotypes from house finches and other songbirds are labelled starting with ‘finch’, followed by the two-letter code for the state in which the haplotype was first isolated, the last two digits of the calendar year in which the haplotype was first isolated, and a single arbitrary number unique to that haplotype. All haplotypes from domestic poultry are labelled with ‘Poultry’, the last two digits of the calendar year of the first isolate identified with the haplotype and then an arbitrary number.

The results that (figure 2b) revealed that differences among the isolates from songbirds were minor in comparison with the variation among M. gallisepticum from poultry. Because of this similarity among songbird isolates, a more appropriate method for describing the associations among the isolates was to use statistical parsimony to create a haplotype network describing the relationships within the recently diverged finch clade of M. gallisepticum (figure 2b). We used TCS 1.21 [33] to construct this haplotype network.

The results that (figure 2b) revealed that differences among the isolates from songbirds were minor in comparison with the variation among M. gallisepticum from poultry. Because of this similarity among songbird isolates, a more appropriate method for describing the associations among the isolates was to use statistical parsimony to create a haplotype network describing the relationships within the recently diverged finch clade of M. gallisepticum (figure 2b). We used TCS 1.21 [33] to construct this haplotype network.
M. gallisepticum typically isolated from house finches and other songbirds, and those present in poultry in the southeastern United States during a similar prolonged time period (figure 2a). Note also the minimal variation among the finch haplotypes relative to that found for M. gallisepticum from poultry (figure 2a). The large number (figure 2b), large geographical extent (figure 1) and long time period (see the electronic supplementary material, table S1) from 1994 to 2011 over which the initial haplotype, ‘Finch_VA94.1’, was isolated, in concert with rapid diversification into multiple lineages (figure 2b), are consistent with a recent colonization and subsequent radiation of M. gallisepticum among house finches [34] accompanied by occasional spillover into other songbirds. The persistence of the initial finch haplotype should not be taken to mean that the genetic make-up of this haplotype has remained unchanged through time, but merely that no changes were present in the marker genes used in this study: full-genome sequences of five isolates from this haplotype (‘VA94’, ‘NC95’, ‘NC96’, ‘NY01’ and ‘WI01’ in figure 4 of [24]) all differed slightly from each other.

Regarding the existence of repeated transfer of M. gallisepticum from poultry to house finches, we have confirmed the suggestion [16] of a second poultry-clade isolate from house finches. This poultry-type haplotype of M. gallisepticum (labelled ‘Finch_NY01.1’ in figure 2a) that was isolated from an asymptomatic house finch in Ithaca, NY in 2001 falls clearly outside the ‘house finch’ clade of M. gallisepticum, as indicated by the high posterior probability score. This second detected transfer to house finches is also consistent with expectation that most host transfers will not result in persistence of the pathogen in a novel host, because this haplotype, isolated from a single house finch displaying no outward clinical signs of disease, was not subsequently detected in songbirds. Unfortunately, we have very few isolates from Ithaca between 2001 and 2011 and examined only three for this study, so we have no effective way of knowing how long the poultry-like haplotype (Finch_NY01.1) persisted in house finches.

We have also identified a new instance in which M. gallisepticum was transferred from songbirds to poultry. This is the second such case, following the one described by Ferguson et al. [18]. We found the ‘Poultry_03.4’ haplotype (black-filled circle in figure 2b) in four isolates from domestic turkeys in North Carolina in 2003. Because these turkeys were from four separate commercial farms belonging to the same company, we believe that they represent a single back-transfer of M. gallisepticum to poultry with subsequent poultry-to-poultry transmission within the company’s facilities. The posterior probability score of 100% overwhelmingly supported placing these isolates with the ‘house finch’ clade of M. gallisepticum. This by itself is not conclusive evidence for a host jump back to poultry from house finches, because an alternative explanation is that a finch-like clade of M. gallisepticum originated and continues to persist in poultry alongside what we have termed the ‘poultry’ clade. This latter explanation would be consistent with two coalescence analyses that date the divergence of M. gallisepticum in house finches from a few years before the first isolation of the bacterium from finches [17,24]. However, our isolate of finch-clade M. gallisepticum from turkeys is intermediate between two isolates from house finches in a network diagram of the finch-clade bacteria (figure 2b). This intermediate position is more parsimoniously explained by a back-transfer from the evolving bacteria in house finches rather than the continued presence in poultry of the lineage containing this haplotype. The viability of a ‘house finch’ M. gallisepticum in domestic poultry is not surprising, given that an isolate of M. gallisepticum from a house finch, probably very similar to the index isolate from house finches, has been experimentally introduced and found to be viable in chickens [35]. Interestingly, the severity of disease that resulted from this experimental infection was so mild that the Finch_VA94.1 strain was suggested as a potential vaccine strain of M. gallisepticum for domestic poultry [36]. Such low virulence in its original host indicates that local adaptation to house finches and concomitant lowered virulence in the original host taxa had occurred already within the first known haplotype of ‘house finch’ M. gallisepticum.

While we have documented only a very small number of transfers of M. gallisepticum between hosts as disparate as poultry and finches, we believe these observations show that transfer rates are high enough that they did not limit the emergence of M. gallisepticum as a pathogen of house finches. Streicker et al. [12] reached a similar conclusion in a study of movement of rabies lineages between bat species, and the ranges of hosts that can be infected by more catholic pathogens such as Salmonella [37] is also consistent with frequent opportunities for transfer of pathogens among host species. While we have documented only a small number of host transfers—confirming a second transfer from poultry to house finches and identifying one new transfer of finch-clade bacteria back to poultry—in probabilistic terms, these events are not rare: one new transfer to house finches from 60 isolates (1.7% of songbird isolates examined), and one new transfer to poultry from 47 isolates (2.1% of poultry isolates). We suggest that these are minimum estimates of transfer rates, because an unknown proportion of transfers will not be detected either because of immediate failure of the bacteria to establish or because the cross-host parasite lineages will have gone extinct rapidly [3,5,6]. Such extinctions could be due either to poor adaptation to the novel host per se, or to poor competitive ability relative to the established ‘finch clade’ of M. gallisepticum. New poultry-clade isolates that cause little or no obvious disease in songbirds will also be under-sampled by our bias towards sampling obviously diseased songbirds. Further, many transfers of house finch M. gallisepticum back to poultry are likely to go undetected, because the poultry most likely to be monitored for infection come from large commercial farms that have biosecurity measures in place to minimize contact between poultry and wild birds, whereas many exchanges may occur with smaller non-commercial (backyard) flocks of poultry with little or no biosecurity.

Our third objective was to determine whether M. gallisepticum in house finches in eastern and western North America represent independent lineages, and whether the relationships among eastern and western M. gallisepticum isolates are consistent with a single colonization of the pathogen from eastern North America house finches [38]. While a previous study indicated genomic divergence of a single western isolate (labelled ‘Finch_CA06.1’ in figure 2) relative to other isolates [24], our phylogeny (figure 2b) indicates that the western isolates we have examined all have a common origin and are derived from the haplotype of M. gallisepticum originally isolated from finches in eastern North America. Our phylogenetic reconstruction gave a posterior probability score of 97% for this conclusion.
...


