Avian diversification across space

van Els, Paul

Chapter V: From pampa to puna: founder events stimulate diversification of a group of Neotropical obligate grassland birds (*Anthus:* Motacillidae)

Diversification of Neotropical *Anthus*

Paul van Els[1,2], Heraldo V. Norambuena[3,4], Rampal S. Etienne[2]

[1]Museum of Natural Science, Department of Biological Sciences, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, U.S.A.

[2]Groningen Institute for Evolutionary Life Sciences, University of Groningen, PO Box 11103, Groningen 9700 CC, The Netherlands

[3]Programa de Doctorado en Sistemática y Biodiversidad, Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Casilla 160-C, Concepción, Chile

[4]Centro de Estudios Agrarios y Ambientales, Casilla 164, Valdivia, Chile.

Abstract

The evolution of Neotropical birds of open landscapes remains largely unstudied. We investigate the diversification and biogeography of a group of Neotropical obligate grassland birds (*Anthus*: Motacillidae). We use a multilocus phylogeny of 22 taxa of *Anthus* to establish a link between diversification and the development of grasslands in South America. We employ the R package DDD to analyze the dynamics of diversification across time in Neotropical grasslands, explicitly testing for shifts in dynamics associated with the Miocene development of C4 grasslands and putative Pleistocene expansion of arid biomes. A lineage-through-time plot revealed increases in the number of lineages, and DDD detected potential shifts to a higher clade-level carrying capacity during both these periods. However, the shifts were non-significant, possibly due to power issues. We used BioGeoBears to investigate ancestral areas, and directionality of colonization of Neotropical grasslands. Members of the genus diversified into, out of, and within the Andes. A DIVALIKE biogeographic model including founder event speciation was most likely. IMa2 analysis revealed that gene flow likely hinders speciation over short to moderate distances, and combined with our biogeographic analysis, we infer that speciation in the group occurs most likely after long-distance dispersal events across forest or elevational zones.

KEYWORDS: Andes, dispersal, Neotropical grasslands, puna, savanna, Pleistocene, Llanos

Introduction

The study of diversification in Neotropical birds has been centered largely on the rich Amazonian and Andean forest biota, as evidenced by an abundance of recent phylogenetic and phylogeographic studies (e.g., Fjeldså, Bowie & Rahbek, 2012; Fernandes *et al.,* 2015; Harvey & Brumfield, 2015). Although forests have provided fruitful foci of study, approximately 15% of South America is covered in various types of natural open lowland and montane grassland (Eva *et al.,* 2004), which hold a unique avifauna of open habitats. The dynamics of diversification in open landscapes may differ greatly from those in forested habitats. For example, grassland taxa are presumably more vagile due to seasonal climatic fluctuations and fire regimes that force movements of grassland inhabitants (Hovick, Elmore & Fuhlendorf, 2013; Little, Hockey & Jansen, 2013). The ability to disperse across the landscape matrix is a major factor in determining rates of diversification in birds (Smith *et al.,* 2014). Increased dispersal capacity may lead to increased gene flow among populations and a reduction of genetic diversity, but may also result in the establishment of new populations across landscape barriers and increase genetic diversity through founder effects ('intermediate dispersal model', Claramunt *et al.,* 2012; Diamond, Gilpin & Mayr, 1976; Phillimore *et al.,* 2006).

The difference in geographic distribution of grasslands versus forest may also produce differences in the spatiality of diversification patterns in grassland versus forest birds. Grasslands and other semi-arid habitats form a geographic complement to forests in much of South America. In the Andes, extensive grasslands occur almost exclusively between 2500-4800 m (Román-Cuesta *et al.,* 2014). Andean grasslands are often distributed in a continuous way along the mountain chain, similar to Andean forests. Unlike Andean forests, they represent the highest vegetation zone, being often highly isolated from each other by intervening lower, forested habitats, that potentially act as barriers to gene flow (Cuervo 2013; Robbins & Nyári 2015). Finally, tropical grasslands, are found in a ring around the Amazon Basin ('circum-Amazonian' distribution, Remsen *et al.,* 1991), with scattered pockets within the Amazon Basin. The unique configuration of Neotropical grassland landscapes likely has a profound influence on the biogeography of grassland organisms and their phylogeography may differ considerably from those of better studied forest taxa.

Grasslands also have very different origins from forests, and this undoubtedly has had profound effects on the evolutionary trajectory of its inhabitants. Similar to other continents, the emergence of grasslands in the Neotropics occurred much later than the rise of forests (Pennington & Hughes 2014). Whereas Neotropical grasslands originated during the Paleogene, C3 cold-adapted grasslands were not widespread until the mid-Cenozoic, and C4 warm-climate grasslands expanded much later, during the late Neogene (Strömberg, 2011). There is some evidence that grassland-inhabiting organisms show an uptick in diversification during the development of grasslands (Estep *et al.,* 2014; Neiswenter & Riddle 2010; Agarwal & Ramakrishnan, 2017), mainly during the Miocene. Temporal estimates of the evolution of tropical C4 grasslands range from the early Miocene (Edwards *et al.,* 2010; Strömberg, 2011) to the late Miocene specifically for South America (Pennington & Hughes 2014). Many bird groups are too young to test for a

correlation between the Miocene spread of grasslands and diversification, having diversified mainly during the Pliocene and Pleistocene, under the influence of climatic fluctuations, especially in temperate and montane areas (Lovette, 2005; Weir, 2006). If the Miocene spread of grasslands plays a role in the diversification of grassland-restricted organisms, diversification should, however, be fastest during or right after the Miocene, with a plateauing of diversity during the Pliocene and Quaternary. Although now often contested (Arruda *et al.,* 2017, Burkart, 1975; Bush & Oliveira, 2006; Colinvaux, 1998; Patel *et al.,* 2011), Haffer (1969) postulated that the distribution of Neotropical lowland taxa was also strongly affected by Pleistocene climate changes. Amazonian forest retreated to refugia, whereas open, drier habitats including grasslands dominated the Amazon Basin. If this holds true, the periods before and after the Last Glacial Maximum that are characterized by a relatively modest extent of dry biome should have left a genetic footprint in populations of grassland organisms that contrasts with that of the LGM. However, this idea has not been explicitly evaluated, due to the lack of available information about the diversification of organisms of the Neotropical arid biome.

To address this issue, we focused on the diversification of a group of Neotropical grassland birds, the Neotropical pipits (*Anthus*), which are represented in the New World by nine species (SACC version 2017, Remsen *et al.,* 2017), five of which are polytypic. All resident New World *Anthus* except for the North American *A. rubescens* Tunstall and *A. cervinus* Pallas form a monophyletic group (Alström *et al.,* 2015; Voelker, 1999a,b; Van Els & Norambuena 2017). Neotropical pipits are obligate grassland birds found in all temperate, tropical, and montane grassland areas in South America, including some of the smaller Amazonian and Andean patches. Hence, this genus is ideal for exploring the timing, rate, geographic direction, and geographic variation of diversification of grassland birds across Neotropical open landscapes.

Some species of *Anthus* are known for regular (Mild & Alström, 2010) and irregular long-distance movements (Voelker, 2001; Brinkhuizen *et al.,* 2010; Lees & VanderWerf, 2011). Voelker assessed dispersal patterns at a broad, mainly intercontinental scale within the entire genus using dispersal-vicariance analysis (DIVA, Voelker, 1999b). Methods have recently become available that rely on likelihood calculation and incorporate several biologically relevant parameters, including founder effect speciation events (Matzke, 2012), to investigate diversification across the landscape, in conjunction with ecological parameters such as dispersal capacity. These models can be combined with time-calibrated phylogenies to more accurately infer the timing and geography of dispersal patterns at a smaller, intra-continental scale. This will not only shed light on the direction of diversification in this group of grassland-dependent birds, but also will allow us to further test the hypothesis that grassland bird diversification in South America is correlated with the spread of dry biomes on the continent.

In brief, the aim of this study is 1) to test whether the timing and rates of diversification of Neotropical *Anthus* broadly agree with the Miocene spread of lowland grasslands, and the Pleistocene spread of Andean and expansion of Amazonian grasslands, 2) to verify what the geographic sequence was of the *Anthus* spread to all major grassland areas on the South American

continent across time, and 3) to investigate if gene flow is a pervasive factor in the population genetics of these taxa.

Materials and methods

Samples and data collection

We obtained 39 tissue samples of all 22 subspecies-level taxa (Table 1, Appendix 1) within the 'New World *Anthus* clade' (Voelker, 1999a; Alström *et al.,* 2015) the remaining New World taxa belong to a mostly Asian clade. Most taxa are represented by >1 individual from localities as widely dispersed as possible within their ranges (Fig. 1). To further investigate phylogeographic patterns in the Neotropical lowlands, we used 66 samples from throughout the range of *A. lutescens* (Appendix 1), the only widespread tropical lowland taxon in the clade. These samples were used in the phylogeographic and population genetic components of this study. Based on existing phylogenies of *Anthus*, we chose *A. cinnamomeus* Rüppell, *A. gustavi* Swinhoe, *A. rubescens,* and *A. rufulus* Vieillot as outgroups, because they represented closely related clades (Alström *et al.,* 2015; Voelker, 1999a). We extracted total genomic DNA from pectoral muscle using a Qiagen DNeasy tissue extraction kit (QIAGEN, Valencia, California) following manufacturer's protocol. In a few cases we used toe pad tissue, for which we first washed toe pads three times with ddH_2O, extended incubation to 24 hours, and added diothiothreitol (DTT) to the incubation stage, extended the elution step to one hour, and eluted twice to a total volume of 300 ul, after which we vacufuged the total volume down to 150 ul.

We amplified one mitochondrial gene (NADH dehydrogenase subunit 2 – ND2) and three nuclear genes: intron 2 of the myoglobin gene MYO (Heslewood *et al.,* 1998), intron 5 of the beta-fibrinogen gene FIB5, and intron 9 of the sex-linked gene for aconitase ACOI9 (Kimball *et al.,* 2009). We used the primer sequences listed in Appendix 1 for amplification of mitochondrial and nuclear genes, and designed several internal primers specific for *Anthus* for amplification of ancient DNA using Geneious 8.1 (Kearse *et al.,* 2012).

We performed polymerase chain reactions (PCR) in 12.5 µl reactions using the following protocol: denaturation at 94 °C for 10 min, 40 cycles of 94 °C for 30 s, variable annealing temperatures, and 72 °C for 2 min, followed by 10 min elongation at 72 °C and 4 °C soak. We used the program Sequencher (Gene Codes Corporation, Ann Arbor, Michigan) for alignment. To detect and interpret insertions and deletions in the nucDNA, we used the program Indelligent (Dmitriev and Rakitov 2008). We phased sequences in DnaSP using the algorithm provided by PHASE (Stephens & Donnelly, 2003), with an ambiguity cutoff of >0.7. We deposited sequences in GenBank (accession numbers listed in Table 1).

Phylogenetic analyses

We identified the best-fit nucleotide substitution model for each locus using jModeltest 2 (Darriba *et al.,* 2012; Guindon & Gascuel, 2003). The HKY+I model was the best fit for all loci. We recovered a species tree in *BEAST, a component of BEAST v. 2.3.2 (Drummond & Rambaut, 2007), achieving ESS values > 200 for all parameter values. We used 'coalescent: constant size' for the tree prior, which is suitable for analyses at relatively shallow phylogenetic levels, and we ran the analysis for 100 million generations, sampling every 1000. We analyzed posterior output in TRACER v. 1.5 and specified a burn-in of 10%. ND2 data were determined to be clocklike in MEGA7.0 (Kumar *et al.,* 2008; AIC = 2692.016). We used a lognormal substitution rate prior with a mean of 2.9×10^{-8} substitutions/site/year (Lerner *et al.,* 2011) for ND2 and nuclear rates of 1.35×10^{-9} substitutions/site/year (Kimball *et al.,* 2009), applying lognormal distributions for most user-specified priors, except for base-frequency proportions (uniform) and population size priors (Jeffrey's). There is no fossil data from South America to verify the timing of diversification of the group, but a fossil *Anthus* from Pliocene deposits (4.3–4.8 mya) from southwestern Kansas shows features that overlap with mean morphometrics of *A. spraguei* (Emslie, 2007) and was used as a minimal age calibration point for this lineage, as well as for the ancestor of Neotropical *Anthus* for a more conservative estimate. To verify the topology estimated in BEAST2, we also constructed an ML tree in Garli 2.0 (Zwickl, 2006), using 1000 bootstrap replicates and the same nucleotide substitution model settings as used for the BEAST2 analysis. Biogeographic and diversification analyses were run on a reduced dataset (n = 17), from which we pruned subspecies-associated lineages that were genetically indistinguishable from others or differed only minimally in previous population-level studies (Van Els & Norambuena 2017).

Diversification analysis

To test for effects of diversity-dependence across our tree, we used the R package DDD v. 3.5 (Etienne *et al.,* 2012; Etienne & Haegeman, 2012), which enables maximum likelihood estimations of diversity-dependent diversification (function: dd_ML), as well as testing for major shifts in these parameters across the tree (function: dd_SR_ML, 'shifting-rates model'). We explicitly tested for the effects of major climatic events on the dynamics of *Anthus* diversification by examining different models with fixed time parameters under a shifting-rates framework: at 5.3 mya (Miocene-Pliocene boundary to test for effects of C4 grassland spread in South America), at 4.0 mya (start of formation of uppermost Andean vegetation zones, Hughes and Eastwood 2006), at 2.0 mya (end of formation of uppermost Andean vegetation zones, Hughes & Eastwood, 2006) and at 1.8 mya (start of Quaternary glaciations). We ensured a full search of parameter space by assessing models using fixed time parameters at every 1-million-year interval in the tree. As a baseline for comparison we also calculated likelihood estimations of a diversity-independent constant-rate birth-death model (function: bd_ML). For model comparison we used AIC and the

bootstrap likelihood ratio test of Etienne et al. 2016 with 10,000 bootstrap replicates. We made a lineage-through-time plot of our data using the R package 'ape'.

Model-based biogeographic analysis

We used the dispersal-extinction-cladogenesis (DEC, Ree & Smith, 2008), dispersal-vicariance-like (DIVALIKE, Ronquist *et al.,* 1997), and Bayesian analysis of biogeography when the number of areas is large (BAYAREA, Landis *et al.,* 2013) models in the package BioGeoBears (Matzke, 2012) implemented in R v.3.2.0. BioGeoBears optimizes ancestral range states onto internal nodes of a tree and produces likelihood estimates of the transitions between states on these nodes. The DEC model implies that transitions between biogeographic regions takes place mainly at cladogenesis events, and allows for the estimation of dispersal, local extinction, and cladogenesis parameters. The DIVALIKE model functions in a similar likelihood framework as the DEC model, but excludes certain biogeographic scenarios including subset sympatry. BAYAREA, finally, only allows for events to happen along branches, and not at cladogenesis events. BioGeoBears also incorporates the parameter *j* into the three abovementioned models, which allows for estimations of the contribution of founder event speciation to the development of the ancestral range of taxa. We constructed a geographical range matrix, coding each taxon as present or absent in one or multiple areas. We included the following geographic regions in the model: Andes, lowlands east of the Andes, lowlands west of the Andes and the area north of the Panamanian Isthmus (including North America). Varying the maximum number of areas a taxon can occupy from 2 to 4 had little effect on likelihood estimates. We did not apply time stratification or distance multipliers, but we ran a separate analysis placing a constraint on adjacency between areas (where east and west of Andes, and north and south of the Panamanian Isthmus were considered non-adjacent).

Phylogeographic and population genetic analyses

We assessed phylogeographic patterns in *A. lutescens* Pucheran, a relatively well-sampled taxon with a wide distribution in the Neotropical lowlands, using mitochondrial data only. For other taxa, we compared uncorrected pairwise genetic differences between samples from geographically distant localities. We produced a Bayesian skyline plot of southern populations of *A. lutescens* (with large n) in BEAST v. 2.3.2. and calculated population statistics in HIERFSTAT (Goudet, 2005) implemented in R v.3.2.0. To estimate gene flow between populations, we used the Isolation-with-Migration (IMa2) software (Hey & Nielsen, 2007). We analyzed the following three pair-wise comparisons of populations based on mtDNA (ND2): northern clade versus southern clade, Suriname versus Guyana, and NW Argentina versus C Bolivia. We were not able to calculate migration rates between certain pairs of subpopulations that extensively shared haplotypes, which violates the minimum amount of divergence required by IMa to produce interpretable parameter distribution values. We applied a substitution rate prior with a mean of 2.9×10^{-8} substitutions/site/year (Lerner *et al.,* 2011) and converted this rate to substitution rate

per locus by multiplying by the length of the ND2 locus expressed in number of nucleotides. We ran ~10 trials to identify appropriately calibrated model parameter priors, after which we used a burn-in period of 500,000 steps followed by 2–2.5 million iterations (>150 effective sample size for each parameter).

Results

Topology

ND2 was represented by 292/1041 variable sites, ACOI9 by 62/960, FIB5 by 37/581, and MYO by 34/723. Our species tree (Fig. 2), obtained using *BEAST, reveals a major split between two groups of *Anthus*: one group consists of small-bodied taxa mostly found in the lowlands (*A. lutescens, A. furcatus* d'Orbigny and Lafresnaye, *A. spraguei* Audubon), with the exception of *A. brevirostris* Taczanowski. The other group contains a mix of Andean and lowland taxa (*A. hellmayri* Hartert, *A. bogotensis* Sclater, *A. correndera* Vieillot, *A. chacoensis* Zimmer, *A. nattereri* Sclater). *A. peruvianus* Nicholson (traditionally a subspecies of *A. lutescens)* is basal to this group. For further discussion of taxonomy, we refer to Van Els & Norambuena (2017).

Timing and diversification

According to our *BEAST2 analysis, the ancestor of New World *Anthus* is estimated to have evolved ~7.5 mya and subsequently diversified mainly on the South American continent (Fig. 3), where the main biogeographic split took place in the early Pliocene between mainly lowland and Andean taxa. More recent speciation events are mainly associated with tips that are associated with an Andean state.The diversity-dependent model (Table 2) with a parameter shift at ~5.8 mya was the most likely model according to AIC, followed by a secondary optimum at ~1.7 mya. Likelihood bootstrap analyses indicate that although our LTT plots and diversification analyses show a shift in dynamics at these points in geological time, the size of our phylogenetic tree likely prevents us from finding statistical support at $\alpha = 0.05$ (Fig. 4).

Biogeography

BioGeoBears revealed DIVALIKE models were most likely (Table 3). BioGeoBears allocates the oldest node in the tree (basal node) with highest likelihood to uncertain geographic origin (Fig. 2). The subsequent splits in the two main clades within the group are most likely between an Andean origin for *A. bogotensis meridae/ A. b. bogotensis/ A. hellmayri* and an eastern lowland origin for the small-bodied subclade, with a preceding split of *A. peruvianus* occurring on the Pacific coast of South America.

The directionality of dispersal patterns in Andean taxa is variable; there is likely at least one lowland-to-Andes dispersal events (*A. furcatus/A. brevirostris*) and two Andes-to-lowland dispersal events (*A. hellmayri brasilianus/A. h. hellmayri* and split between lowland and Andean *A. correndera*). Several taxa speciated within the Andes; the first split occurred between the northern Andean *A. b. meridae* and the other taxa, followed by diversification between the northern and relatively isolated southern Andes (*A. b. shiptoni, A. h. dabbenei*).

Phylogeographic patterns and gene flow

Within continuous landscapes haplotype diversity is relatively low, and individuals sampled from opposite ends of the continuous geographical distribution (> 500 km apart) of taxa often exhibit minimal genetic variation (Appendix 1), ranging from 0% sequence divergence in *A. h. hellmayri, A. l. lutescens*, and *A. peruvianus* to 0.5% in the montane *A. bogotensis*. Phylogeographic patterns in *A. lutescens* (Fig. 5) reveal a dominance of a single haplotype in the southern part of its distribution. North of Amazonia, every isolated grassland block in our study system is represented by a unique haplotype (three savannas in Guyana, one in Suriname, and one in Panama). Northern populations have roughly twice the haplotype diversity of southern populations (Appendix 1), with the exception of Corrientes, Argentina, which has a comparable level of haplotype diversity to northern populations. *Anthus peruvianus* is represented by four samples from three widely scattered localities throughout its range, all with identical haplotypes.

The Bayesian skyline analysis (Fig. 6) reveals a steadily expanding population towards the present from about 0.5 mya. Tajima's D and Fu's F_s were negative and significant for all southern populations together (but not for any individual population), indicating possible population expansion or purifying selection. F_{st} values are negative between Guyana and other populations, confirming that Guyana harbors relatively high genetic variation. Southern populations, with the exception of Tucumán, are characterized by low F_{st} values (Appendix 1). Our gene flow analysis indicated nearly non-existent levels of gene flow between populations north and south of the Amazon river (m_1, N->S = 0.16±0.18, m_2, S->N = 0.09±0.14, with population mutation parameters q_1 = 20.34±10.84, q_2 = 47.04±12.28), and higher effective population sizes south (N_e = 47.04) than north (N_e =20.34) of the Amazon. Gene flow between populations north of the Amazon River was higher than across the Amazon, from Suriname (q_1 = 205.89±296.40) to Guyana (m_1 = 4.99±2.91) and vice versa (q_2=499.86±288.75, m_2 = 4.67±2.96), as was gene flow between populations south of the Amazon, from C Bolivia (q_1 = 9.57±8.69) to NW Argentina (m_1 = 2.26±2.25) and vice versa (q_2 = 25.47±14.37, m_2 = 2.73±2.60).

Discussion

Early diversification: Miocene-Pliocene

The sister lineage to the New World *Anthus* is Eurasian (Voelker, 1999a,b; Alström *et al.,* 2015), Voelker (1999a) suggested that dispersal occurred via the Bering Strait. *A. spraguei* is the only North American member of the clade, and this taxon likely dispersed from South to North America (in agreement with Voelker, 1999b). As far as we are aware, no fossils of *Anthus* are known from Central America. Many Mexican grasslands date from the Tertiary (Rzedowski, 1975) and it is curious that the Mexican and Central American grasslands (north of Panama) are devoid of resident *Anthus*, especially because taxa with similar ecologies such as *Cistothorus* (Robbins & Nyári, 2015) and *Sturnella* (Barker, Vandergon & Lanyon, 2008) occur throughout the region. The paucity of North and Central American *Anthus* pertaining to our New World clade could indicate that a long-distance dispersal event from Asia to South America, rather than arrival via a Bering Sea crossing, resulted in the colonization of the New World by *Anthus.* This hypothesis is not as far-fetched as it may seem, given that that long-distance dispersal from Asia to South America is known to occur in *Anthus* (Brinkhuizen *et al.,* 2010). Long-distance dispersal and subsequent diversification is known in Motacillidae from other regions as well; Alström *et al.* (2015) suggest that two members of the family established populations in their isolated African and Wallacean winter quarters and evolved into morphologically and ecologically highly divergent taxa.

Given our phylogeny, the Neotropical ancestral *Anthus* has unknown geographic origins within the New World. The topology may however provide some insight into the evolutionary origins of the two major Neotropical subclades of *Anthus.* The first major split possibly arose due to an ancestral *Anthus* being isolated on either side of the Andes or between the Andes and the eastern lowlands. Some of the deepest branches in our tree are represented by taxa not included in Voelker (1999a,b). These deep branches have a relatively large potential to subdivide shorter branches or groups and are thus comparatively informative (Wiens, 2006). Indeed, our recovery of phylogenetic relationships, the estimation of timing of diversification and the determination of dispersal patterns in Neotropical *Anthus* has been enabled by the sampling of relatively rare taxa that remained un-sampled in earlier decades.

Our LTT plot shows a distinct increase in lineages at the end of the Miocene and beginning of Pliocene, followed by a leveling out of diversification. Diversification analysis showed that this increase may be associated with the initial post-colonization radiation of *Anthus* lineages across available New World grassland habitats. However, likelihood ratio testing showed that there are no significant diversity-dependent effects associated with the Miocene spread of grasslands, possibly as a result of small sample size.

Recent diversification: Quaternary

We furthermore find evidence for a possible change in diversity dynamics between ~1.7-1.8 mya, which suffers from the same power issues as the Miocene shift (data not shown, but similar to that in Fig. 4c,d). This change does not however suggest an increase of lowland lineages in line with a Pleistocene expansion of lowland dry biomes according to the refuge hypothesis (Haffer, 1969). Only two lineages occur exclusively in the tropical lowlands, which may be caused by ecological or physiological limits to the expansion of *Anthus* into tropical lowland grasslands, but are confirmed by population genetic data. Populations of *A. lutescens* remained relatively stable until about 0.5 mya, and then increased steadily towards the present with no sign of higher rates during the Pleistocene.

Most of the new lineages responsible for the uptick in diversification towards the present are Andean, rather than lowland, in origin. Although we lack exact data on the development of Andean habitats in relation to orogeny, it seems clear that large expansions of the highest vegetation zones probably developed during the last ~4.5 mya of most intense orogeny (Hoorn *et al.,* 2010; Vuilleumier, 1965), and continue to the present. Our lineage-through-time plot shows a period of stasis in diversification during the Pliocene, followed by a gradual accumulation of lineages around the start of the Pleistocene. This contrasts with relatively low Quaternary diversification rates in other birds (Zink *et al.,* 2004; Zink & Klicka, 2006), as well as with diversification patterns in other grassland-specialist organisms (Neiswenter & Riddle, 2010; Agarwal & Ramakrishnan, 2017). However, unlike in our study, Pleistocene bursts in diversification may in part have been missed due to incomplete sampling of genetic lineages embedded within sampled species (often based on more conservative taxonomies). The continuing expansion of grasslands in the still-rising Andes may contribute to these patterns. Although *Anthus* has mostly colonized the Andes before the Pleistocene, most of the within-Andes diversification seems to have occurred during the last 2 mya. Many recent diversification events are also Andean in origin in other Neotropical grassland-inhabiting organisms such as *Cistothorus* wrens (Robbins & Nyári, 2015), *Hypericum* St. John's worts (Nürk *et al.,* 2013), and *Lupinus* lupines (Hughes & Eastwood, 2006). The formation of Andean 'islands' with the completion of isolated high peaks likely contributed to the diversification of these taxa (Cuervo, 2013).

Diversification north and south of the Amazon

Anthus is represented in lowland South America north of the Amazon by only one lineage (*A. l. parvus* (see Van Els & Norambuena, 2017 on taxonomy). During the last 1 mya a split occurred between populations of *A. lutescens* north and south of the Amazon Basin. While several landscape-level processes may have led to this split, it is likely that the transformation of the South American landscape from a 'cratonic' (based on geologically stable shield-formations) to an Andean-dominated system (Hoorn *et al.,* 2010) with correlated climatic changes, contributed to landscape changes in and around the Amazon Basin that led to isolation of both groups. A number

of other bird species have a similar distribution pattern north and south of the Amazon (circum-Amazonian distribution, a similar pattern found in many African savanna birds, distributed around the Congo Basin), and further investigation should reveal if 1. northern populations are generally derived from southern populations, and if 2. the timing of the split is similar to that found in *A. lutescens.*

Taxa occurring in grasslands north of the Amazon are generally a subset of those occurring in southern South America, with few endemic representatives (Stotz *et al.,* 1996; e.g. Tyrannidae, Thraupidae). While the relatively modest extent of grasslands north of the Amazon relative to southern South America may explain this pattern to some extent, other factors likely contributed to this pattern as well. Northern South American grasslands are probably much younger than their southern counterparts (and were colonized much later according to our biogeographic analyses), resulting in a discrepancy in diversity. Regardless of the mechanism driving these diversity patterns, southern South America seems to be a center of diversification for Neotropical lowland grassland birds.

Supporting the north-south split in *A. lutescens* in our tree, a large mitochondrial sample of *A. lutescens* from south and north of the Amazon Basin revealed a dichotomy between the two areas; the north harboring relatively many unique haplotypes. Northern populations are found in savanna islands in a sea of forest, and gene flow analysis indicates there is exchange of genetic material between these populations. Southern grasslands are more continuous in distribution, facilitating dispersal and gene flow between populations. Furthermore, the presence of an individual with a southern haplotype in Suriname indicates 1. a possible vagrant, 2. a possible migration pattern, 3. recent incursion of southern birds into Suriname. We consider the possibility of sampling a vagrant unlikely, but we cannot rule out any option with certainty. The distance between the sampling site and the nearest population of the southern clade is ~500 km. Considerable deforestation has taken place north of the Amazon River between Santarém and the Sipaliwini savanna in the last 50 years (Skole & Tucker, 1993), and the newly created patches of open habitat may provide a corridor for the two clades to enter into contact.

Dispersal promotes and reduces diversification

Sequence divergence found in geographically remote samples (all > 500 km) in several continuously distributed Neotropical *Anthus* taxa is similar to that found in a similar system of cerrado birds, with samples 1800 km apart (Bates, Tello & Silva, 2003). Divergence was much lower than that in Amazonian birds (Bates *et al.,* 2003; Nyári, 2007; Sousa-Neves *et al.,* 2013) found on opposite sides of a river.

The relative lack of structure in *Anthus* even over large distances in continuous habitats is not just attributable to short diversification times and is likely the result of widespread gene flow within continuous habitats. What exactly the mechanisms are behind gene flow patterns will remain to be discovered, but likely are related to relatively high dispersal capacity and life history

characteristics inherent to life in the grasslands: the tracking of seasonal resources and burns, and the crashing and breaking up of populations in response to these factors from year to year. Over longer distances, the effect of dispersal and consequent gene flow disappears. Isolated grasslands are represented by unique genetic lineages. Our biogeographic models indicate that founder event speciation is important in the group, indicating that dispersal is a factor that may enhance gene flow over short distances or in continuous habitats, but that it is simultaneously driving diversification over larger distances and in isolated habitat patches. DIVALIKE models of biogeography were chosen over DEC, which in cases where a taxon can occupy a maximum of four areas simultaneously, is most likely due to a lack of subset sympatry. This means that evolution of sister species is not likely to happen in the same geographic area, another testament to the fact that gene flow may hinder diversification over shorter distances or in the absence of considerable vicariant effects.

In summary, the initial diversification of South American Anthus was driven by Andean-lowland vicariance, and later diversification occurred mainly by founder events into, out of, and within the Andes. Our lineage-through-time plot shows upticks in diversification during both the end of the Miocene and the Pleistocene. However, we did not detect statistically significant shifts in diversification parameters, likely due to small sample size of our tree. Gene flow is pervasive within continuous stretches of grassland, leading to little diversification without the aid of major vicariant barriers, coupled with dispersal.

Acknowledgments

We thank the following institutions and their staff for providing samples: Paul Sweet (AMNH), Nate Rice (ANSP), Stephen Massam (FIMNT), John Bates and Ben Marks (FMNH), Krzysztof Zyskowski (YPM), Mark Robbins (KU Biodiversity Institute), Sharon Birks and John Klicka (UWBM), Brian Schmidt (USNM, Smithsonian), Jon Fjeldså and Jan Bolding Kristensen (ZMUC), Jeremiah Trimble (MCZ), Kimball Garrett (LACM), Pablo Tubaro and Darío Lijtmaer (MACN), Alexandre Aleixo (MPEG), Jorge Pérez-Emán at the Instituto de Zoología Tropical at the Universidad Central de Venezuela, and Pedro Victoriano at the Universidad de Concepción in Chile. Andy Wood at the British Antarctic Survey provided valuable samples of *A. c. antarcticus*. Sabrina Taylor at LSU's Department of Renewable Natural Resources kindly allowed me to work in her ancient DNA lab. James V. Remsen, Jr. and Robb Brumfield kindly reviewed the manuscript. Funding was provided by the LSU Museum of Natural Science Birdathon Fund, the P.A. Hens Memorial Fund for Systematics, an American Ornithologists' Union Alexander Wetmore Memorial Research Award, and the American Museum of Natural History Frank M. Chapman Memorial Fund. PVE thanks the Faculty of Science and Engineering and the Groningen Institute for Evolutionary Life Sciences at the University of Groningen for funding through the Adaptive Life Program. HVN is grateful for the CONICYT-PCHA/DoctoradoNacional/2013-21130354 scholarship. RSE thanks the Netherlands Organization for Scientific Research (NWO) for financial support through a VICI grant.

Table 1. Taxon sample list, including institution, tissue number, country, region and Genbank accession number per locus. Asterisks denote sequences obtained from historical samples. Institution codes are as follows: AMNH, American Museum of Natural History; BAS, British Antarctic Survey; FIMNT, Falkland Islands Museum and National Trust; KU, University of Kansas Natural History Museum; KUSNM, Danish Natural History Museum at University of Copenhagen; LSUMZ, Louisiana State University Museum of Natural Science; MCZ, Museum of Comparative Zoology at Harvard; UCCC, Universidad de Concepción; USNM, Smithsonian Institution National Museum of Natural History; UWBM, University of Washington Burke Museum; and YPM, Yale Peabody Museum.

Taxon	Institution	Tissue	Country	Region	ND2	MB	FGB5	ACOI9
antarcticus	BAS	2	South Georgia	-	MF320010	MF320015	MF320056	MF320047
antarcticus	BAS	3	South Georgia	-	MF320009	MF320016	MF320057	MF320048
bogotensis	KUSNM	116859	Ecuador	Cotopaxi	MF319979	MF320095	MF320070	MF320027
bogotensis	LSUMZ	431	Peru	Piura	MF320026	MF320094	MF320069	MF320026
immaculatus	KU	25127	Peru	Ayacucho	MF320028	MF320105	MF320074	MF320028
*meridae**	AMNH	811977	Venezuela	Mérida	MF320011	-	-	-
*meridae**	AMNH	811978	Venezuela	Mérida	MF320012	-	-	-
shiptoni	USNM	645734	Argentina	Tucumán	MF320000	MF320111	MF320080	MF320034
shiptoni	UWBM	54394	Argentina	Tucumán	MF319999	MF320110	MF320079	MF320033
*chacoensis**	AMNH	797085	Argentina	Córdoba	MF320008	-	-	
calcaratus	LSUMZ	61430	Peru	Puno	MF319985	MF320084	MF320051	MF320016
calcaratus	LSUMZ	61431	Peru	Puno	MF319986	MF320085	MF320052	MF320017
catamarcae	UWBM	54511	Argentina	Tucumán	MF320001	MF320012	MF320081	MF320044
chilensis	AMNH	13589	Argentina	Río Negro	MF320035	MF320100	MF320060	MF320035
chilensis	AMNH	13591	Argentina	Río Negro	MF320036	MF320101	MF320061	MF320036
correndera	USNM	630116	Uruguay	Tacuarembó	MF319989	MF320088	MF320055	MF320020
grayi	FIMNT		Malvinas/Falklands	-	MF320007	MF320102	MF320071	MF320037
brevirostris	KU	21673	Peru	Puno	MF319996	MF320103	MF320072	MF320038
brevirostris	KU	21681	Peru	Puno	MF319997	MF320104	MF320073	MF320039
furcatus	UWBM	54556	Argentina	Tucumán	MF347705	MF320113	MF320082	MF320045
furcatus	USNM	635884	Uruguay	Artigas	MF320002	MF320114	MF320083	MF320046
brasilianus	UWBM	54574	Argentina	Corrientes	MF319991	MF320090	MF320059	MF320022
brasilianus	USNM	630210	Uruguay	Tacuarembo	MF319990	MF320089	MF320058	MF320021
dabbenei	UCCC	2376	Chile	Araucania	MF320013	MF320117	-	MF320049
dabbenei	UCCC	2377	Chile	Araucania	MF320014	MF320118	-	MF320050

hellmayri	KU	9813	Argentina	Jujuy	MF319994	MF320108	MF320077	MF320042
hellmayri	UWBM	54528	Argentina	Tucumán	MF319995	MF320109	MF320078	MF320043
abariensis	USNM	626029	Guyana	Parabara	MF319987	MF320086	MF320053	MF320018
abariensis	YPM	13701	Suriname	Sipaliwini	MF319988	MF320087	MF320054	MF320019
lutescens	LSUMZ	87109	Bolivia	Santa Cruz	MF320003	MF320098	MF320067	MF320029
lutescens	USNM	645602	Argentina	Tucumán	MF320004	MF320099	MF320068	MF320030
parvus	LSUMZ	41613	Panama	Bocas del Toro	MF319982	MF320093	MF320064	MF320025
peruvianus	LSUMZ	44804	Peru	La Libertad	MF319984	MF320097	MF320066	MF320032
peruvianus	LSUMZ	48218	Peru	Lima	MF319983	MF320096	MF320065	MF320031
nattereri	KU	3604	Paraguay	Itapúa	MF319992	MF320106	MF320075	MF320040
nattereri	KU	3665	Paraguay	Itapúa	MF319993	MF320107	MF320076	MF320041
spraguei	LSUMZ	25702	U.S.A.	North Dakota	MF319980	MF320091	MF320062	MF320023
spraguei	LSUMZ	21749	U.S.A.	Louisiana	MF319981	MF320092	MF320063	MF320024
cinnamomeus	UWBM	52816	South Africa	Eastern Cape	AY329410	-	-	-
gustavi	UWBM	75556	Russia	Primorsky Krai	HM538396	-	-	-
rubescens	LSU	53141	U.S.A.	California	MF320015	-	-	-
rufulus	FMNH	358350	Philippines	Sibuyan	KP671566	-	-	-

Table 2. Output of tests for diversity-dependent diversification in package DDD and protracted speciation in package PBD. CR=constant-rates, DD=diversity-dependent, SR=shifting-rates (number indicates time of shift), k = number of parameters, max log lik = maximum log likelihood of model, AICw = Akaike Information Criterion weights, λ = speciation rate, μ = extinction rate, K = clade-level carrying capacity (after 1. first and 2. second shift in case of a SR model), tshift = time at which shift occurs in diversity dynamics (in case of SR model).

model	k	max log lik	AICw	λ	μ	K1	K2	tshift
CR	2	-34.697	0.000	0.269	0.001	-	-	-
DD	3	-31.504	0.021	2.047	0.324	16.577	-	-
SR	5	-28.377	0.065	1.157	0.001	8.999	19.489	1.682
$SR_{5.8}$	5	-25.874	0.791	9.542	0.344	1.273	16.386	5.856
$SR_{5.3}$	4	-31.257	0.009	1.639	0.326	94.784	16.613	5.300
$SR_{4.0}$	4	-31.381	0.009	2.211	0.301	13.855	16.516	4.000
$SR_{2.0}$	4	-31.141	0.011	1.549	0.156	11.544	17.171	2.000
$SR_{1.8}$	4	-29.021	0.092	1.097	0.001	8.999	19.641	1.800

Table 3. Results ancestral range estimation analyses from BioGeoBears, using no constraints on adjacency of four defined biogeographic areas (first six rows), and using constraints on adjacency between areas east and west of Andes, and between north and south of Panamanian Isthmus (last six rows) df is degrees of freedom per model, LnL is log likelihood, AICw is weight of Akaike Information Criterion, d is dispersal, e is extinction, and j is founder effect speciation.

Model	df	-LnL	AICw	d	e	j
DEC	2	-30.724	0.001	0.032	0.035	0.000
DEC+j	3	-19.545	0.166	0.000	0.000	0.142
DIVALIKE	2	-25.942	0.001	0.031	0.000	0.000
DIVALIKE+j	3	-18.685	0.393	0.000	0.000	0.125
BAYAREAS	2	-38.794	0.001	0.033	0.222	0.000
BAYAREAS+j	3	-20.675	0.054	0.000	0.000	0.148
DECc	2	-21.080	0.097	0.120	2.005	0.000
DECc+j	3	-19.507	0.173	0.089	3.420	0.053
DIVALIKEc	2	-25.942	0.001	0.030	0.000	0.000
DIVALIKEc+j	3	-21.405	0.026	0.051	0.504	0.108
BAYAREASc	2	-38.733	0.001	0.064	0.095	0.000
BAYAREASc+j	3	-20.165	0.090	0.075	1.821	0.065

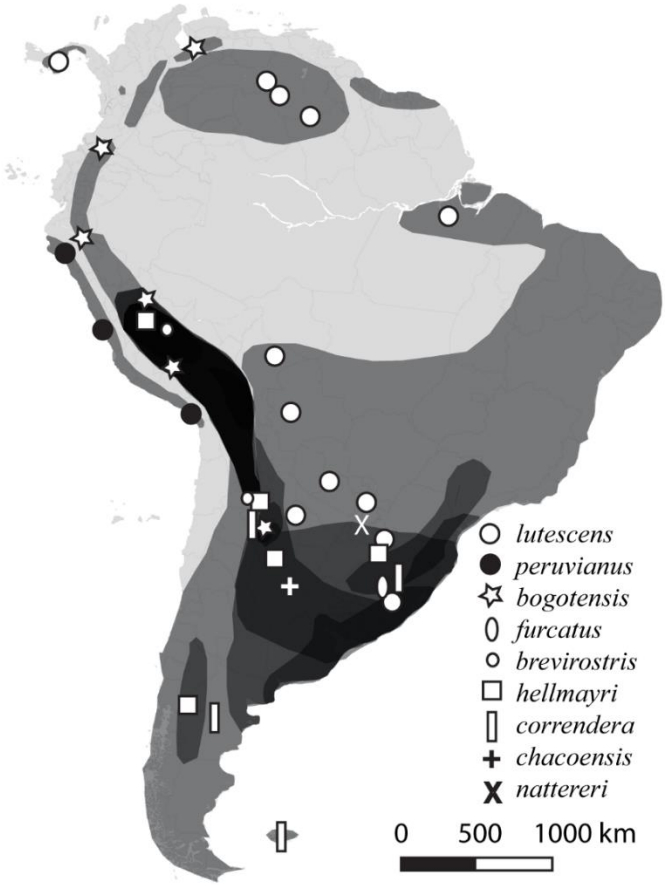

Figure 1. Sampling map of Neotropical *Anthus*, *A. lutescens* (white circles), *A. peruvianus* (black circles), *A. bogotensis* (stars), *A. furcatus* and *A. brevirostris* (ovals), *A. hellmayri* (squares), *A. correndera* (rectangles), *A. chacoensis* (hatch), *A. nattereri* (cross). Background shade represents taxon diversity, darker being more taxa.

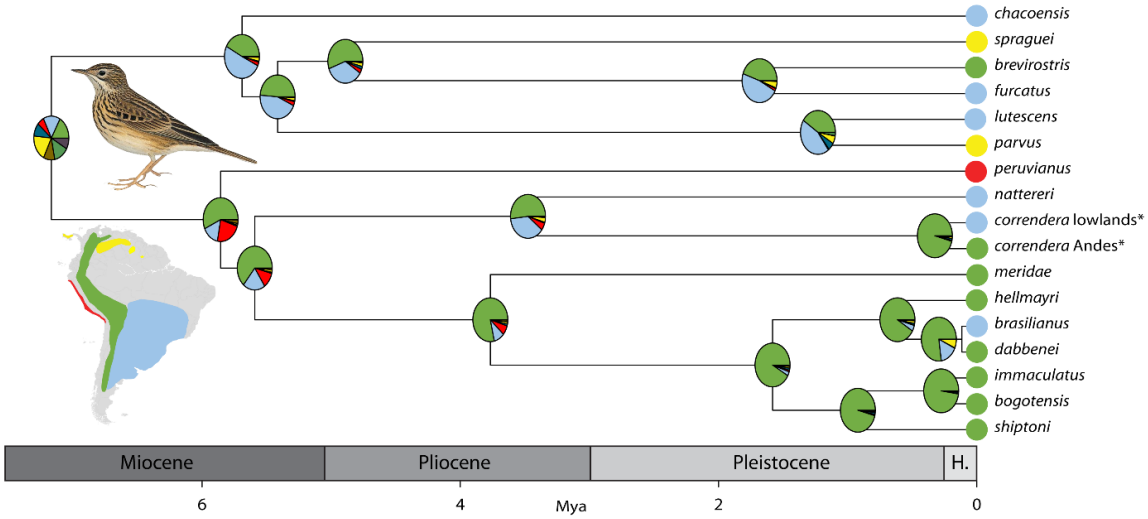

Figure 2. Biogeography and diversification of Neotropical *Anthus*. For support values see Van Els and Norambuena 2017. Pie charts indicate ancestral range states at each node according to DIVALIKE+j model in BioGeoBears: blue lowlands south of the Amazon Basin and east of Andes, green is Andes, yellow is lowlands north of Amazon Basin and area north of Isthmus of Panama, red represents the Peruvian coastal strip, other colors are uncertain states (also see inset map). Values at bottom of tree indicate divergence time in millions of years. **Correndera* lowland is made up of *A. c. correndera/chilensis/grayi/antarcticus*, *correndera* Andes is *A. c. catamarcae/calcaratus*. Pipit illustration Tyler 2004.

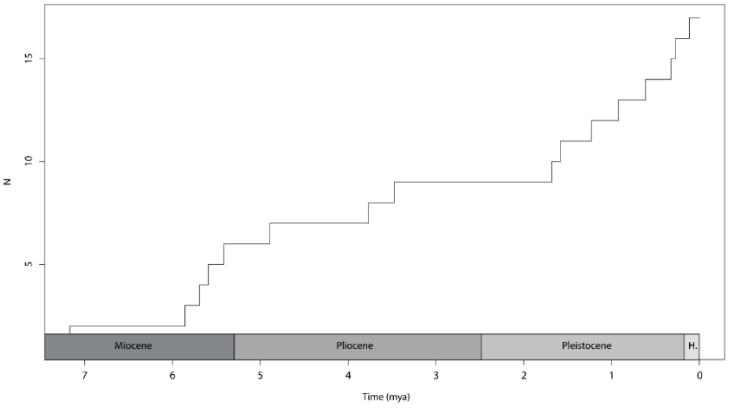

Figure 3 Lineages-through-time plot of the Neotropical *Anthus* clade. Black line is the number of lineages, minus subspecies-level lineages in *A. correndera* (cf. Figure 2). Shaded area represents 95% credibility interval. Dashed lines from left to right represent late Miocene/Pliocene boundary and associated completion of spread of lowland grasslands, lower bound on uppermost Andean vegetation zones, upper bound on uppermost Andean vegetation zones, and the start of Quaternary glaciations.

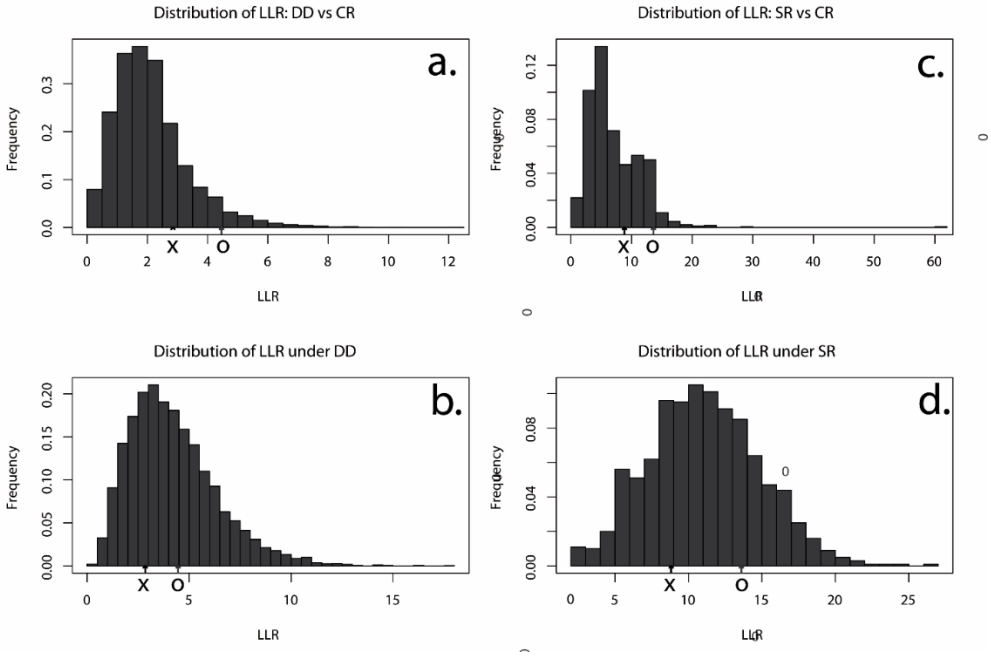

Figure 4. Bootstrap likelihood ratio test for Anthus diversification analysis in DDD. Distribution of logarithms of the likelihood ratio generated in DDD under an a. constant-rates (CR) model for diversity-dependence, a b. diversity-dependent model (DD), c. a constant-rates model for shifting-rates, and d. a shifting rates (SR) model using the parameters of the best SR model in Table 2 to test for effects of diversity-dependence. Crosses represent values of the likelihood ratio for the real data, and circles represent values of the likelihood ratio for a significance level of α=0.05 with 10,000 replicates.

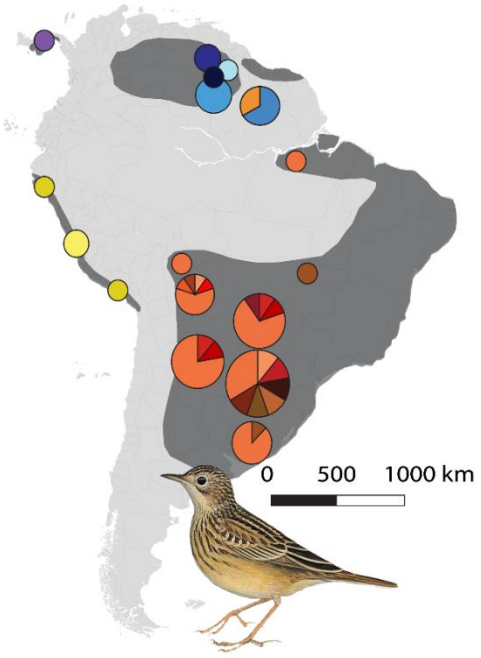

Figure 5. Mitochondrial haplotypes of *Anthus lutescens* and *A. peruvianus*. Size of circles represents sample size, colour indicates clade (blue are *parvus*, red are *lutescens*, yellow are *peruvianus*) and hue represents haplotype.

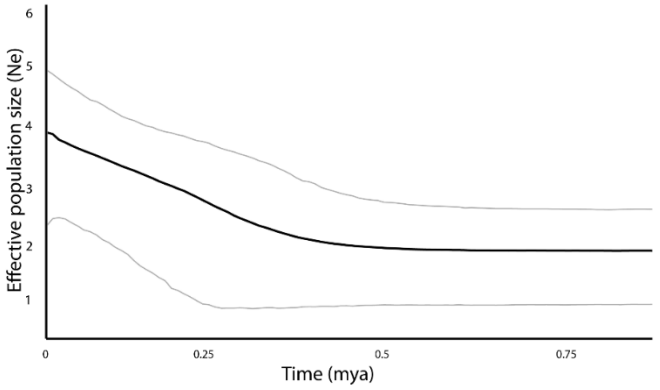

Figure 6. Historic population size of *A. lutescens lutescens*, based on ND2 gene of 53 individuals. Black line represents median population size, grey lines represent upper and lower bounds of 95% posterior density interval.

Table S1. *Anthus lutescens* sample list, including institution, tissue number, country, and region.

Taxon	Institution	Tissue	Country	Region
lutescens parvus	USNM	622349	Guyana	Parabara
lutescens parvus	USNM	626029	Guyana	Parabara
lutescens parvus	USNM	622350	Guyana	Parabara
lutescens parvus	USNM	625794	Guyana	Parabara
lutescens parvus	USNM	622269	Guyana	Parabara
lutescens parvus	USNM	625978	Guyana	Karaudanawa
lutescens parvus	USNM	627497	Guyana	Kusad Mtn.
lutescens parvus	YPM	13701	Suriname	Sipaliwini
lutescens parvus	YPM	2255	Suriname	Sipaliwini
lutescens parvus	USNM	632436	Guyana	Kopinang
lutescens parvus	USNM	632640	Guyana	Kopinang
lutescens parvus	USNM	626238	Guyana	Wiwitau
lutescens parvus	YPM	2255	Suriname	Sipaliwini
lutescens parvus	YPM	13701	Suriname	Sipaliwini
lutescens parvus	LSUMZ	41613	Panama	Bocas del T.
lutescens lutescens	LSUMZ	87109	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	38007	Bolivia	Santa Cruz
lutescens lutescens	USNM	645602	Argentina	Tucumán
lutescens lutescens	USNM	636034	Uruguay	Rocha
lutescens lutescens	USNM	645601	Argentina	Tucumán
lutescens lutescens	USNM	635798	Uruguay	Rocha
lutescens lutescens	USNM	636034	Uruguay	Rocha
lutescens lutescens	USNM	636033	Uruguay	Rocha
lutescens lutescens	USNM	645602	Argentina	Tucumán
lutescens lutescens	USNM	645313	Argentina	Tucumán
lutescens lutescens	USNM	636108	Uruguay	Rocha
lutescens lutescens	USNM	645314	Argentina	Tucumán
lutescens lutescens	USNM	635833	Uruguay	Rocha
lutescens lutescens	UWBM	54570	Argentina	Corrientes
lutescens lutescens	UWBM	54571	Argentina	Corrientes
lutescens lutescens	UWBM	54495	Argentina	Tucumán
lutescens lutescens	UWBM	54567	Argentina	Corrientes
lutescens lutescens	UWBM	54580	Argentina	Tucumán
lutescens lutescens	UWBM	54581	Argentina	Tucumán
lutescens lutescens	UWBM	54572	Argentina	Corrientes
lutescens lutescens	UWBM	70446	Argentina	Tucumán
lutescens lutescens	UWBM	54500	Argentina	Tucumán
lutescens lutescens	UWBM	54568	Argentina	Corrientes
lutescens lutescens	UWBM	54499	Argentina	Tucumán
lutescens lutescens	UWBM	54998	Argentina	Tucumán
lutescens lutescens	UWBM	54569	Argentina	Corrientes
lutescens lutescens	UWBM	54494	Argentina	Tucumán
lutescens lutescens	UWBM	54496	Argentina	Tucumán
lutescens lutescens	UWBM	54497	Argentina	Tucumán
lutescens lutescens	UWBM	70771	Argentina	Corrientes
lutescens lutescens	UWBM	92247	Argentina	Tucumán
lutescens lutescens	UWBM	70523	Argentina	Corrientes
lutescens lutescens	UWBM	54575	Argentina	Corrientes

lutescens lutescens	UWBM	54576	Argentina	Corrientes
lutescens lutescens	KU	3789	Paraguay	Itapua
lutescens lutescens	KU	3249	Paraguay	Pres. Hayes
lutescens lutescens	KU	3245	Paraguay	Pres. Hayes
lutescens lutescens	KU	3333	Paraguay	Ñeembucú
lutescens lutescens	KU	3334	Paraguay	Ñeembucú
lutescens lutescens	KU	3320	Paraguay	Misiones
lutescens lutescens	KU	3293	Paraguay	Misiones
lutescens lutescens	LSUMZ	77761	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	39771	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	9628	Bolivia	Pando
lutescens lutescens	LSUMZ	79763	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	38053	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	38054	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	38035	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	38004	Bolivia	Santa Cruz
lutescens lutescens	LSUMZ	48432	Bolivia	Santa Cruz
lutescens lutescens	YPM	137257	Suriname	Sipaliwini