Epidemiology and ecology of virus and host: bats and coronaviruses in Ghana, West Africa

Heather Joan Baldwin
BA, BSc (Hons 1)

Department of Biological Sciences
Faculty of Science and Engineering
Macquarie University
and
Institute of Evolutionary Ecology and Conservation Genomics
Ulm University
by cotutelle

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

2015

MACQUARIE University
SYDNEY·AUSTRALIA

universität ulm

For Elisabeth Kalko, who inspired all who knew her,
and my parents, Robert and Judith Baldwin, whose unending support
and encouragement made it possible for me to follow my passion.

Image: Marco Tschapka

TABLE OF CONTENTS

ABSTRACT

Bats are implicated in the emergence of many zoonotic diseases, including the coronaviruses (CoVs) responsible for deadly outbreaks of severe acute respiratory syndrome (SARS) in 2002-2003, and Middle East respiratory syndrome (MERS) which first emerged in 2012. Despite an increased interest in bat CoVs since the SARS epidemic, little research has examined bat CoVs from a disease ecology perspective. In this thesis, I investigated the epidemiology and ecology of CoVs and bats that host them in Ghana, West Africa. I captured bats at several colonies regularly over two years. Samples were screened for CoVs, which were found to be widespread in insectivorous cave-dwelling bats of the genera *Hipposideros* and *Nycteris*. A novel CoV related to MERS-CoV was detected at high prevalence in slit-faced bats, *Nycteris* cf. *gambiensis*, supporting suggestions of a bat origin of MERS-CoV. I examined population and individual-level risk factors for infection with four CoVs in *Nycteris* cf. *gambiensis*, *Hipposideros abae*, and *H.* cf. *ruber*. I found a strong seasonal effect on CoV infection rates, and a strong association with age, whereby juvenile bats are at greater risk of infection, as well as evidence for higher risk with low body condition. These findings provide new insights into the ecological, demographic and temporal processes that influence CoV infection dynamics in bats, with implications for public health management to prevent virus spillover to humans and domestic animals.

In the second half of my thesis, I examined the evolution of cryptic species, and genetic and acoustic diversity in CoV-hosts *Hipposideros caffer* and *H. ruber.* Using mitochondrial DNA and microsatellite loci, and acoustic and morphological data, I showed that mitochondrial DNA clades represent distinct species, with no evidence of interbreeding between groups. I then used mitochondrial sequence data to date lineage divergence, and explore the historical and evolutionary processes that may have given rise to diversity in these taxa. Further, I characterised spatial patterns of intraspecific genetic diversity in the three species corresponding to *H.* cf. *ruber* (*H.* sp. B, C, and D). Comparison shows stark contrasts in the spatial patterns of genetic variation of these species, despite them having diverged relatively recently and apparently having highly similar ecological and natural history traits. I discuss the implications of these results for

disease surveillance and spread, wildlife management, and the evolutionary processes that generate and maintain diversity.

The findings of this thesis add to knowledge of bat CoV ecology and epidemiology, as well as to the evolution and patterns of diversity within a relatively widespread, but understudied, group of Afrotropical bats that are natural CoV hosts. The results are discussed in the context of zoonoses, wildlife management and disease prevention, and the evolution of cryptic species in bats.

DECLARATION

I declare that the work presented in this thesis is my own and was undertaken during my PhD candidature. Wherever sources of information and the work of others have been used, I have acknowledged this in the text. This thesis will be submitted to Macquarie University as part of the requirements for the degree of PhD and the Universität Ulm for the degree of *Dr. rer. nat.* according to the joint doctoral supervision agreement (cotutelle).

Heather Joan Baldwin

August 2015

STATEMENT OF CONTRIBUTION

All data chapters are co-authored manuscripts. My contributions to each chapter are outlined below.

	Chapter 2	**Chapter 3**	**Chapter 4**	**Chapter 5**	**Chapter 6**
Reference	Annan *et al.* 2013[§]	Baldwin *et al.* submitted	Baldwin *et al.* 2014	Chapter 5	Chapter 6
Design	30%	95%	95%	90%	90%
Sample collection	70%	85%	100%	85%	85%
Lab work*	10%	100%	95%	95%	95%
Data analysis	30%	100%	100%	90%	90%
Manuscript preparation	20%	95%	95%	90%	90%

* excludes viral diagnostics, which were performed by co-authors and technical assistants acknowledged in each chapter

[§] Joint first authored paper

ACKNOWLEDGEMENTS

There are a great number of people who have accompanied me on this journey, and who have helped and supported me in various ways. I am deeply thankful to all of you.

First and foremost, I would like to thank my exemplary supervisors, Adam Stow, Marco Tschapka, and Elisabeth Kalko, whose guidance, expertise, and mentorship have been invaluable.

Adam, I ended up in your group kind of by accident six years ago, and I'm so happy it worked out that way. Thank you for everything, for the great ideas, solid advice, moral support and friendship, and for your ability to recognise 'that face' and know when I needed a chat. I can't adequately express how truly grateful I am for your support throughout the years. I couldn't have asked for more from a supervisor and mentor.

Marco, ich danke Dir von ganzem Herzen. Thank you for taking on the always interesting (in more ways than one) Ghana project, and for your continued support and guidance through the ups and downs of this journey. Thanks for having and open door and an ear for me when I needed it.

To Eli I am deeply indebted, for giving me the opportunity to join this wonderful team, her support and intellectual guidance, and for her enthusiasm, which was truly infectious. This project and my involvement in it would not have been possible without her, and I am so grateful to have had the opportunity to know and work with her, and for the marvellous time we shared in the field. I wish we could have seen our plans come to fruition together. Leider bist Du viel zu früh gegangen.

I have been incredibly lucky to have had the opportunity to work on three continents with many amazing people as part of this PhD. There are many people whose valuable contributions and collaborations were essential to the research, or whose support in other ways has been vital. Thank you to the KCCR Bat Team, for your camaraderie and companionship on many long nights in the field! Evans Nkrumah, Priscilla Anti, Ebenezer Badu, me nnamfo, I will never forget those times. My friends and colleagues Anna Vogeler, Luci Kirkpatrick, Isaac Mawusi Adanyeguh and Julia Morrison, thanks for giving your all in those hard-core samplings,

and for your support, friendship and humour. Thank you to all of our field assistants and volunteers; Emmanuel Essoun, Kenneth Quansah, Eunice Okyere, Emmanuel Asare, Julia Morrison, Sarah Koschnicke, Michael Owusu, David Ofori Agyei, Elikem Nutsuakor, Justice Konadu, Mr Paul, Kwame Takyi, Winfred, Samuel, Gaddafi, Benewaa, Jesse, and Desmond. Medaase!

This work could not have been undertaken without the support and hospitality of Buoyem, Kwamang, Forikrom, Likpe Todome and Akpafu Todzi communities. I especially thank Nana Ben, Nana Adams, Madame Grace, Nana Emmanuel, Francis, the Chief of Akpafu, Aunty Jane, the helpful staff at Elmina Castle, particularly Mr Arhin, Peter Obin at the University of Cape Coast, and folks at the Ghana Wildlife Division.

I also thank the staff of the Kumasi Centre for Collaborative Research in Tropical Medicine (KCCR), in particular Ingrid Sobel, Kerstin Schand, Augustina Annan, Dr Ellis Owusu-Dabo and Stephen Kwarteng. Thank you to Augustina, Michael Owusu, Richard and several other KCCR staff who performed many CoV detection assays. Ingrid and Kerstin, vielen lieben Dank an ihr beiden für alles. From KNUST I thank Prof Samuel Oppong and Prof Olivia Agbenyega. Peter Sothmann and Clemens Frank, thank you for your friendship, and for sharing in adventures and copious bottles of Gulder! I also thank Thomas Kruppa, Thomas van Kampen, and my Kumasi families, Paul and Patience Marfo and the Abdul-Rahmans. I met too many wonderful people during my time at KCCR to name everyone, but I'm grateful to all for making that time joyful.

I am grateful to have had the opportunity to learn from the outstanding virology group at Bonn University. Thank you to Christian Drosten and Felix Drexler for giving me the opportunity to spend time in your labs. I thank Victor Corman, Tobias Bleicker, Monika Eschbach-Bludau, Sebastian Brünink, Andrea Rasche and the other good folks at Bonn for tirelessly performing an epic number of extractions and assays for the Ghana CoV project, and for your kindness and patience during my time there. Andrea, thanks for sharing your desk, and

Tabea Binger, big love and thanks to you chica for sharing your home, couch, clothes and pretty much everything!

I am grateful to my friends and colleagues in Ulm, for welcoming me and for making my time there so enjoyable. Thank you to Simone Sommer, Manfred Ayasse and Bio3 as a whole. Simone, thanks for showing a keen interest in my work and for helpful advice. Any thanks for technical and administrative support provided by Sebastian Tschunke, Ulrike Stehle, Gabi Wiest-Danner and Ingrid Dillon. Peter Vallo, thanks for the support, advice, and camaraderie. Mirjam Knörnschild and Kirsten Jung for valuable advice and support. Big love to Tania Gonzalez, Tona Ruiz, Maria Eckenweber, Monica Cuervo, Maria Helbig, Julian Schmid, Karina Montero, Tina Mumm, Kirstin Übernickel, Pablo Santos, Mark Gillingham, Simon Ripperger, Patrick Cvecko, Ralph Simon, Marco Mello, Carlos Martel, and Jan Bechler. Anna Vogeler, I can't think of my time in Ulm without reminiscing on our fun times, thank you for the many good memories. Stefan Brändel, thank you for many good talks and for making me laugh, and remember that I am waiting for your visit along with the good people of Cranberry. Inga Geipel, thank you for your friendship and moral support, and for being there for me always, it meant more than you know. Praise Seitan! To Toshi, thanks for the good chats and delicious food! Thank you also to Bärbel and Benni Abt, Ilse Walter, Julia Jedelhauser, Gerhard Klein, and Christoph Herrmann, for many good times in our respective WGs.

To my wonderful Macquarie family! The Conservation Genetics Lab, Siobhan Dennison, Paolo Momigliano, Shannon Smith, Jake Coates, Ben Ofori, Maria Asmyhr, Jono Davis, Elayna Truszewski, Jessica Thompson, Rafael de Fraga, Vincenzo Repaci, Stephen Hoggard, Miranda Christopher, thank you for the moral support, advice, for sharing your expertise, and funny distractions. Siobhan, thanks for your love, encouragement and dear friendship. Heartfelt thanks to Andy Beattie and Christine Turnbull for your mentorship and advice. Marcela Diaz, thank you for your unwavering support and friendship; knowing I can pick up the phone and talk to you about anything, no matter where we are and how long it's been since our last chat, means a lot (heaps!). Big thanks to Tim Pearson, Liette Vandine, Mike Gillings, Kate Umbers, and

Mariella Herberstein, for support and encouragement, and Veronica Peralta, Marie Howitt, Laura McMillan and Sharyon O'Donnell for all of your help and patience.

I gratefully acknowledge the funding bodies who made this research possible. The majority of funding for this project came from the German Research Foundation. Funding for research and travel were also provided by Macquarie University's Higher Degree Research Funding, Postgraduate Research Fund and Macquarie International, and Ulm University's Institute for Evolutionary Ecology and Conservation Genomics.

Last, but by no means least, I wholeheartedly thank my family and friends for their support and love. To my partner André, your companionship, encouragement and patience have been incredibly important to me. Thanks for having faith in me, for being my rock, and above all, thanks for putting up with me! To my parents, Judith and Robert Baldwin, thank you for fostering my interest in science and learning and the natural world, and for the unconditional support you've given me in everything that I do. I can't thank you enough for the opportunities you worked hard to provide me, and I could not have done this without you. To my brothers, James and Christopher, thanks for being there for me with moral (and technical) support, it means a lot to know that you guys have always got my back; to my sisters-in-law Sharon and Charlotte, and my extended family Wendy, Vic, Betty, Robert, Vivienne, Andrew, Duncan, Sarka, Greg, and Kinda, thank you for your keen interest in my endeavours. My niece Luella, thanks for inspiring me and bringing me much joy, especially in the last few stressful months. Much love and gratitude to my dear friends, Tiffany Williamson, Sonia Buckley, Rebekah Doran, and Jane Gribble, for being there in times of crisis and celebration, and for the ample wine therewith.

This has been an amazing journey; challenging, formative, rewarding, and above all, lots of fun! Thanks again to everyone who was part of it!

CHAPTER 1
General introduction

Image: Marco Tschapka

Emerging infectious diseases, zoonoses, and bats

Emerging infectious diseases (EIDs) pose a significant threat to public health and the conservation of biodiversity (Daszak *et al.* 2000; Jones *et al.* 2008; Smith *et al.* 2009). EIDs are defined as diseases that have recently appeared in a new population or host species, or are increasing in incidence, geographic range or severity (Daszak *et al.* 2000). Recent analyses have suggested that the incidence of EIDs is increasing (Jones *et al.* 2008). This is often linked to anthropogenic factors, such as biodiversity loss and ecological change (Keesing *et al.* 2010; Patz *et al.* 2004), misuse of antibiotics (Morse 1995), and the increase in global air travel (Daszak *et al.* 2000).

Zoonoses, diseases transmitted from a non-human animal host to a human host, comprise 60% of known human pathogens (Jones *et al.* 2008). Of those, more than 70% are known to have originated in wildlife species (Jones *et al.* 2008). In recent years, bats (order Chiroptera) have become a focus of international interest due to their role in the emergence and transmission of zoonotic viruses. Bats are proposed as natural reservoirs for viruses causing serious human disease, including severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), Ebola, Marburg, Hendra, Nipah, and rabies (Calisher *et al.* 2006). This propensity to host zoonotic viruses may be linked to chiropteran adaptations including high levels of interspecific roosting sympatry, sociality and gregariousness, longevity, and flight, as well as genetic and functional differences in the immune system (Baker *et al.* 2013; Calisher *et al.* 2006; Dobson 2005; Luis *et al.* 2013; Zhang *et al.* 2013). Indeed, a recent meta-analysis found that bats are host to more zoonotic viruses per species than rodents, another group often implicated in zoonotic disease (Luis *et al.* 2013).

However, despite the recent increase of interest in bat viruses, most of this research has focussed on discovering and describing new viruses and virus diversity in bats (e.g. Corman *et al.* 2013; Tong *et al.* 2009; Woo *et al.* 2007). Relatively little work has focussed on the ecological mechanisms that drive virus emergence and infection dynamics (Drosten 2013). Host ecology,

through component aspects such as phenology or life history, movement patterns, genetic diversity, species interactions, and social behaviour, may influence virus emergence and infection dynamics through generating differences in prevalence and susceptibility, disease spread or affecting transmission patterns, or influencing host switching (Archie *et al.* 2009; McDonald *et al.* 2008; Plowright *et al.* 2015). Therefore, knowledge of host ecology is vital for making predictions about epidemics, and informed management decisions for both public health and wildlife conservation.

Without ecologically informed approaches to management decisions, reactionary policies or actions that aim to decrease the threat of zoonotic disease to humans or domestic animals may be at best ineffective, and at worse, harmful and counter-productive. For example, despite widespread culling of vampire bats (*Desmodus rotundus*) as a strategy to control rabies in Latin America, recent research has shown that such strategies do not reduce rabies seroprevalence (Streicker *et al.* 2012). Streicker *et al.* (2012) found that culling campaigns may actually increase rabies prevalence, as a result of preferentially removing adults, resulting in populations comprised of a higher proportion of juveniles and sub-adults, which are more susceptible and potentially more important for rabies transmission. Similarly, research on long-term badger culling programs in the UK designed to control the spread of bovine tuberculosis has shown that culling influences spatial organization by disrupting territorial behaviour, increasing contact rates with other badgers and cattle, and is associated with increase incidence in some areas (Donnelly *et al.* 2006; Woodroffe *et al.* 2006).

Disease ecology is a relatively new discipline focussed on investigating the interaction of the behaviour and ecology of hosts with that of the biology and dynamics of pathogens. It merges ecology, epidemiology, microbiology, genetics, and modelling to better understand how ecological processes and natural history affect the interaction between host and pathogen. Knowledge gained through such approaches may improve models and allow better forecasting of

risks and infection patterns and spread, with the ultimate goal to predict and prevent emerging infectious disease.

Coronaviruses

Coronaviruses (order *Nidovirales*, family *Coronaviridae*, CoVs) are enveloped, single stranded, positive-sense RNA viruses, which infect a wide variety of vertebrate species (Woo *et al.* 2009). CoVs are classified into four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* (Drexler *et al.* 2014). *Alpha-* and *Betacoronavirus* are associated with infection in mammals, whereas *Delta-* and *Gammacoronavirus* are found mainly in birds (Woo *et al.* 2009). Coronaviruses are known to infect the respiratory, gastrointestinal, or central nervous systems (Vijaykrishna *et al.* 2007). Since the 1960s, two human coronaviruses (hCoV-OC43 and –229E) have been known, causing mild respiratory disease (Saif 2004). However, in 2002–2003, an outbreak of severe acute respiratory syndrome (SARS) resulting in more than 8000 infections and 800 deaths was caused by a novel *Betacoronavirus*, SARS-CoV (Drexler *et al.* 2014; Drosten *et al.* 2003). A likely natural animal reservoir for SARS-CoV was identified in rhinolophid bats (Li *et al.* 2005). In late 2012, another highly pathogenic human CoV, causing Middle East respiratory syndrome (MERS), emerged in Saudi Arabia (Zaki *et al.* 2012), and outbreaks are ongoing in the Middle East and South East Asia.

In the aftermath of the SARS epidemic, several human CoVs, and numerous novel bat CoVs were described (Drexler *et al.* 2014). Recently, CoVs related to human viruses, including SARS, and hCoV-229E, have been identified in bats in West Africa (Pfefferle *et al.* 2009). However, despite the zoonotic potential of these viruses and the high level of human contact with bats in West Africa (Anti *et al.* 2015), the infection dynamics and risk factors in CoV infection bats have yet to be examined.

Bats – cryptic and diverse mammals

Bats are the second most speciose mammalian order, constituting more than 1300 mammalian species (Fenton and Simmons 2015). Bats are incredibly ecologically important, providing numerous ecosystem services (Kunz *et al.* 2011). Frugivorous bats are vital to many forest ecosystems because of their role in pollination and seed dispersal, without which many forest plant communities would not persist (Kunz *et al.* 2011). Insectivorous bats are of enormous economic importance, providing pest control estimated to be worth between US$3.7 - 53 billion to the agricultural industry in North America alone (Boyles *et al.* 2011).

Approximately one quarter of all bat species are threatened with extinction (Mickleburgh *et al.* 2002). Key threatening processes include habitat destruction and fragmentation, as well as conflict with humans, which can be exacerbated by fears related to zoonotic disease. In North America, bats face a new threat from a devastating emerging fungal disease, white nose syndrome (*Pseudogymnoascus destructans*), causing mass population declines and localised extinctions (Cryan *et al.* 2013).

As a result of their relatively cryptic nature, there are many bat species for which we lack basic biological and ecological information. Further, a tendency towards morphological conservation as a result of the consistent selective pressures associated with flight and echolocation, means that cryptic diversity may be more common in bats than other taxa (Jones and Barlow 2004). Bats are, therefore, an interesting group within which to explore the evolution of cryptic species and patterns of genetic diversity, and the underlying processes. Further, accurate species recognition is valuable for identifying zoonotic disease reservoirs and using disease ecology to inform prediction and prevention of emerging infectious diseases.

Study sites

Sub-Saharan Africa is home to more than 100 bat species (Monadjem *et al.* 2010). In West Africa direct and indirect human contact with bats is common (Anti *et al.* 2015). Bats are hunted for

bushmeat, guano is collected from bat caves, as is water for household use and even consumption, and caves that hold spiritual significance are visited for ritualistic purposes, sometimes involving several days of sleeping, eating and praying inside the caves. We observed evidence of many of these behaviours at our sampling sites (figure 1). Additionally, changes in land use and urbanisation in Africa has lead to large-scale degradation and fragmentation of habitats. All of these factors raise the potential of contact between disease reservoirs and humans. Practices such as bushmeat consumption and cave visitation for resource extraction have been linked to serious human viral outbreaks on the African continent (Amman *et al.* 2012; Leendertz *et al.* 2015; Leroy *et al.* 2009; Swanepoel *et al.* 2007), possibly including the recent catastrophic Ebola outbreak in West Africa (Saéz *et al.* 2015).

Thesis overview

In the chapters that follow I investigate ecology and CoV epidemiology in bats in West Africa. This thesis has four proximate aims: 1) to investigate the prevalence of CoVs in Ghanaian cave-dwelling bats; 2) to elucidate risk factors in CoV infection in bats; 3) to resolve taxonomic uncertainty in CoV-hosts in the *Hipposideros caffer* complex, and 4) to explore patterns of genetic and acoustic diversity in the *H. caffer* group in West Africa. The results are used to address broader questions about zoonoses, as well as evolutionary history, patterns and processes, and wildlife management, specifically relating to: 1) host-pathogen dynamics, including the seasonal, demographic and ecological processes that influence CoV infection in bats; 2) the processes that generate and maintain genetic and acoustic diversity; and 3) concordance of patterns of intraspecific genetic diversity among species with similar ecological and life-history traits. Each chapter is a self-contained manuscript for publication in a scientific journal and is formatted according to the specific journal's requirements, with the exceptions that referencing styles have been altered for consistency.

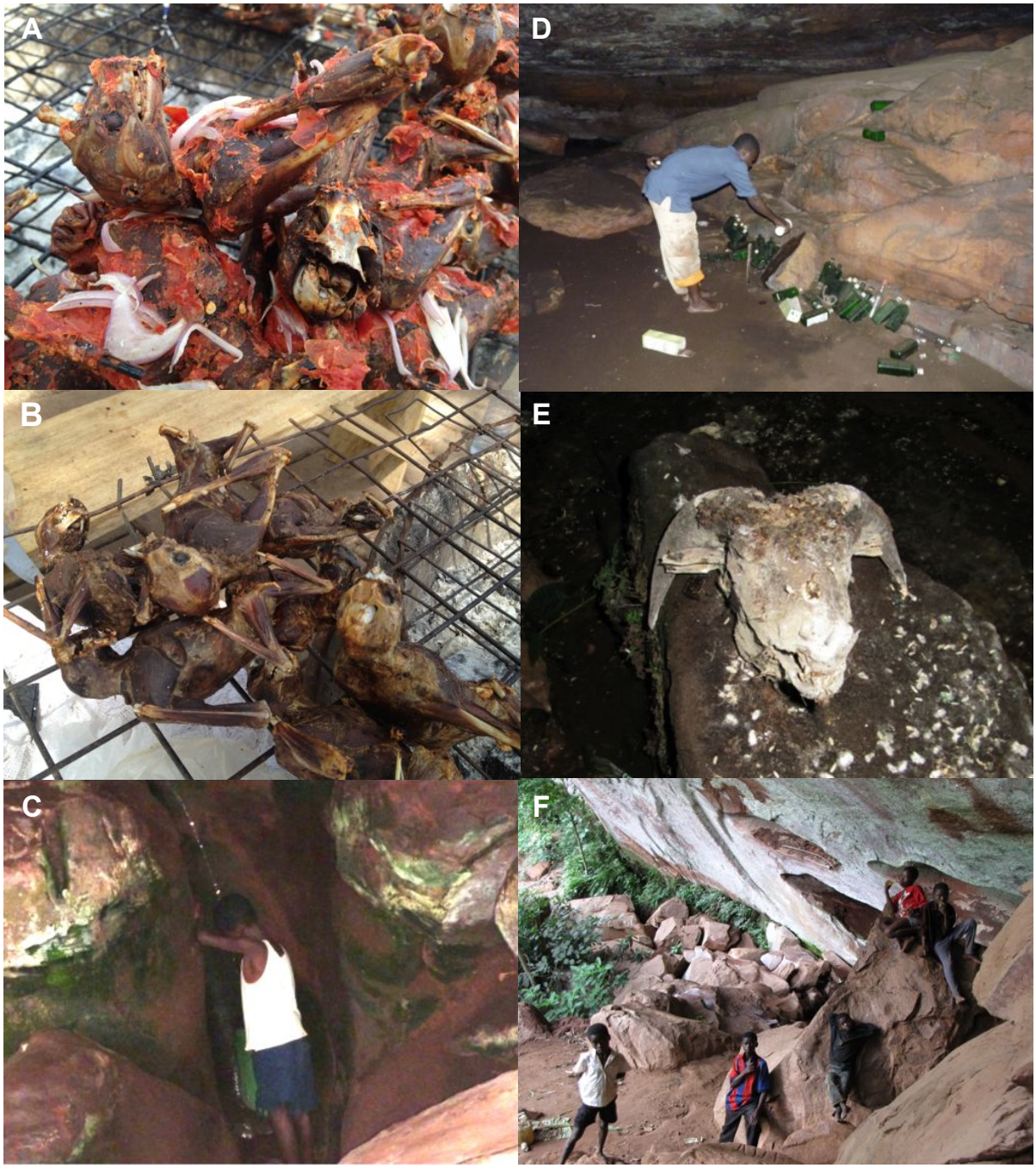


Figure 1. Examples of human contact with bats and cave use in Ghana. a) and b) fruit bats (*Eidolon helvum*) served at a roadside stall in Kumasi, Ghana; c) child collecting water from a stream flowing through Ohene Abutia cave, Kwamang d) a man pouring libation, a religious activity at Mprisi cave in Buoyem. A large number of discarded bottles from previous offerings can be seen; e) a sacrificial animal offering at the Mframmabuom cave in Kwamang; f) children at Ohene Abutia cave, Kwamang, during a water-fetching visit.

Chapter 2 is a publication from early in my candidature. It describes the discovery of CoVs closely related to MERS-CoV (formerly hCoV-EMC/2012) in Ghanaian and European bats. This paper was highly collaborative. It aimed to provide insights into the origins of MERS-CoV following its first emergence in September 2012. This paper was published in *Emerging Infectious Diseases*, under the title "Human betacoronavirus 2c EMC/2012–related viruses in bats, Ghana and Europe".

Chapter 3 describes the results of the first longitudinal study of CoV infection in bats in West Africa. I reported detection and prevalence of four CoVs in multiple bat species from two years of regular surveillance (figure 2). Further, I investigated ecological and demographic risk factors CoV infection in bats in Ghana. This paper has been submitted to *Ecography* and is currently in review.

Previous research (Pfefferle *et al.* 2009) and research presented in Chapter 3 identified *Hipposideros* cf. *ruber* and *H. caffer tephrus*, in the *Hipposideros caffer* complex (figure 3), as being common hosts for CoVs. In order to resolve questions regarding speciation of cryptic taxa in this group and to investigate patterns of genetic diversity and dispersal, I developed a set of microsatellite markers in Chapter 4. This marker set establishes a toolkit for assessing neutral genetic variation in these bats, traditionally recognised as two species, *H. caffer* and *H. ruber.* This paper was published in *BMC Research Notes.*

Chapters 5 and 6 utilise the microsatellite loci developed in Chapter 4 for taxonomic investigation and examination of finescale genetic structuring and connectivity between colonies, and spatial genetic and acoustic variation. In Chapter 5, I resolve some of the taxonomic uncertainty in this group, using molecular, acoustic and morphological data. I date divergences between lineages, and explore the historical and evolutionary processes that may have given rise to diversity in these taxa. This chapter is formatted for submission to the journal *Molecular Biology and Evolution*. In Chapter 6, I compare spatial genetic patterns between sympatric sister species to test the assumption that closely related species with similar traits have responses to extrinsic

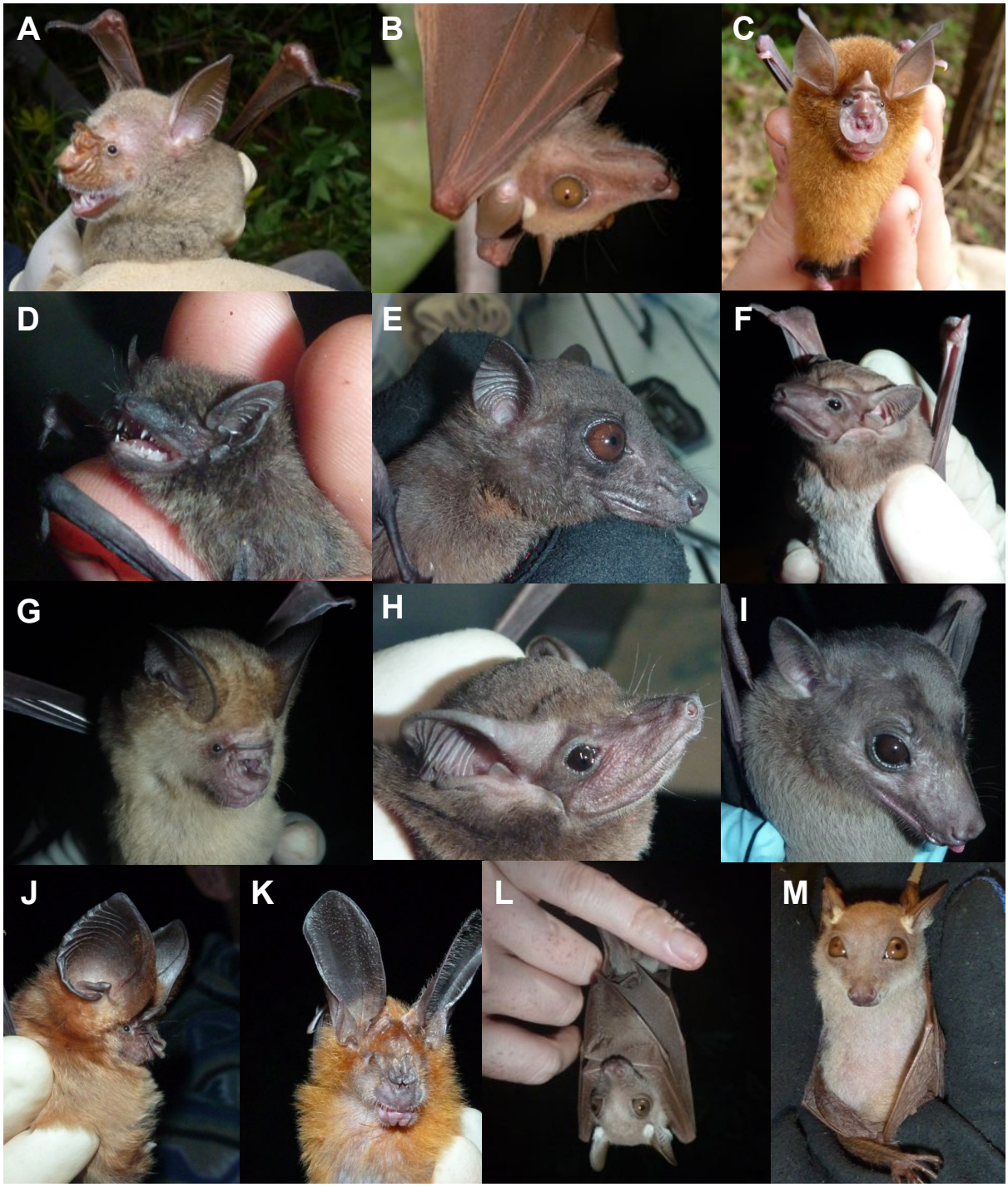


Figure 2. Some of the bat species sampled throughout the fieldwork for this thesis. a) *Hipposideros gigas*, b) *Epomops buettikoferi*, c) *Rhinolophus landeri*, d) pipistrelloid bat, e) *Lissonycteris angolensis*, f) *Taphozous perforates*, g) *H. abae*, h) *Coleura afra*, i) *Rousettus aegyptiacus*, j) *H. jonesi*, k) *Nycteris* cf. *gambiensis*, l) *Micropteropus pusillus*, m) *Nanonycteris veldkampi*

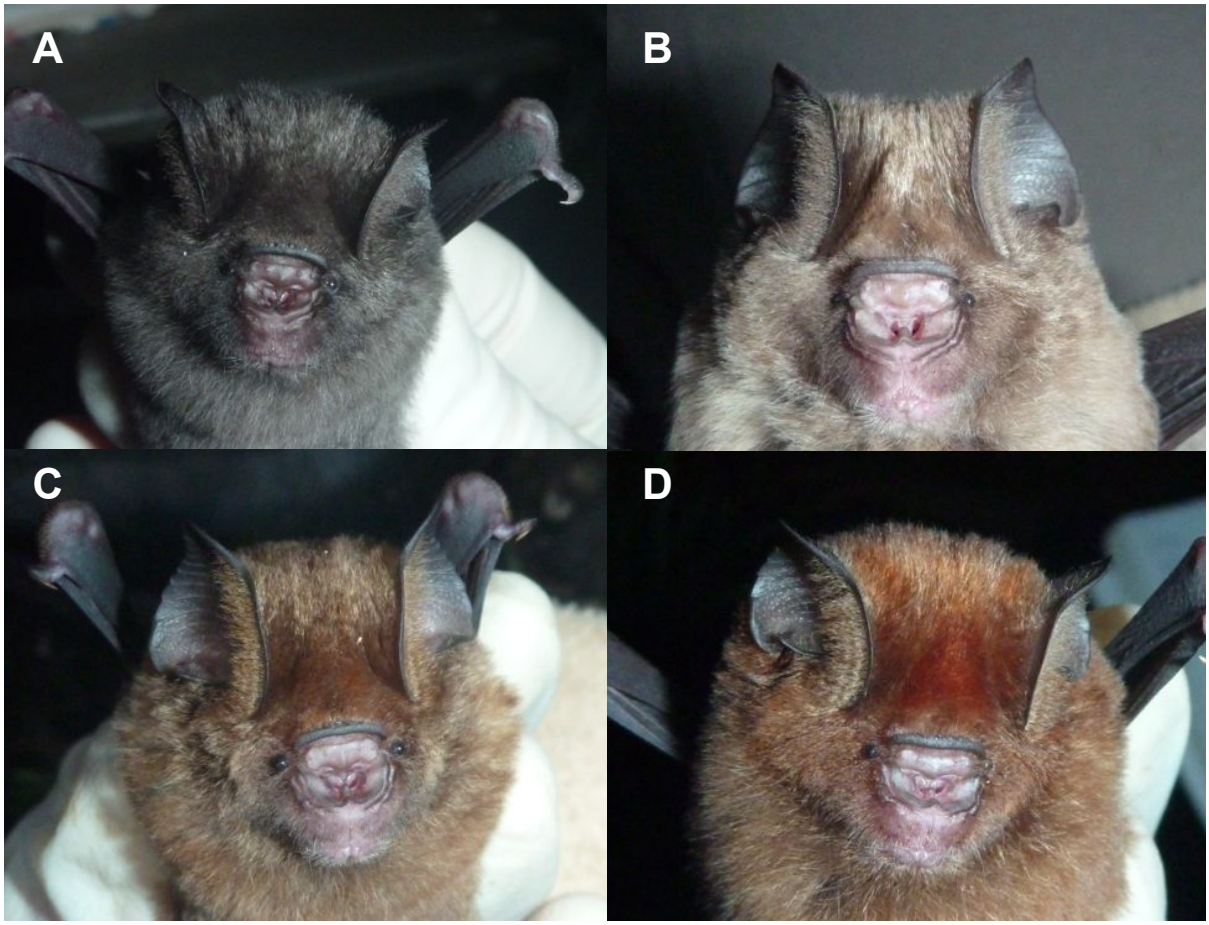


Figure 3. Morphological variation in the *Hipposideros caffer* complex in Ghana. a) *H. caffer tephrus*; b), c) and d) *H.* cf. *ruber*

forces. I reflect on the implications of the findings for wildlife management. This paper is formatted for submission to *Molecular Ecology*.

Chapter 7 synthesises the main findings of this thesis. The Appendix presents other papers authored and published during my PhD candidature. The first of these describes evidence for a bat origin of a human coronavirus, human CoV-229E, and has been published in the *Journal of Virology*. The second deviates from the themes of this thesis, exploring the high prevalence of deceptive strategies in Australian flora and fauna, and was published in *Behavioral Ecology*.

REFERENCES

Amman B. R., Carroll S. A., Reed Z. D., Sealy T. K., Balinandi S., Swanepoel R., Kemp A., Erickson B. R., Comer J. A. & Campbell S. (2012) Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. *PLOS Pathogens* 8(10), e1002877.

Anti P., Owusu M., Agbenyega O., Annan A., Badu E. K., Nkrumah E. E., Oppong S. K., Adu-Sarkodie Y. & Drosten C. (2015) Human–bat interactions in rural West Africa. *Emerging Infectious Diseases* 21(8), 1418-1421.

Archie E., Luikart G. & Ezenwa V. (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution* 24(1), 21-30.

Baker M. L., Schountz T. & Wang L. F. (2013) Antiviral immune responses of bats: a review. *Zoonoses and Public Health* 60(1), 104-116.

Boyles J. G., Cryan P. M., McCracken G. F. & Kunz T. H. (2011) Economic importance of bats in agriculture. *Science* 332(6025), 41-42.

Calisher C. H., Childs J. E., Field H. E., Holmes K. V. & Schountz T. (2006) Bats: important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews* 19(3), 531-545.

Corman V. M., Rasche A., Diallo T. D., Cottontail V. M., Stöcker A., Souza B. F. d. C. D., Corrêa J. I., Carneiro A. J. B., Franke C. R., Nagy M., Metz M., Knörnschild M., Kalko E. K. V., Ghanem S. J., Morales K. D. S., Salsamendi E., Spínola M., Herrler G., Voigt C. C., Tschapka M., Drosten C. & Drexler J. F. (2013) Highly diversified coronaviruses in neotropical bats. *Journal of General Virology* 94, 1984-1994.

Cryan P. M., Meteyer C. U., Boyles J. G. & Blehert D. S. (2013) White-nose syndrome in bats: illuminating the darkness. *BMC Biology* 11(1), 47.

Daszak P., Cunningham A. & Hyatt A. (2000) Emerging infectious diseases of wildlife- threats to biodiversity and human health. *Science* 287(5452), 443-449.

Dobson A. P. (2005) What links bats to emerging infectious diseases? *Science* 310(5748), 628-629.

Donnelly C. A., Woodroffe R., Cox D. R., Bourne F. J., Cheeseman C. L., Clifton-Hadley R. S., Wei G., Gettinby G., Gilks P., Jenkins H., Johnston W. T., Le Fevre A. M., McInerney J. P. & Morrison W. I. (2006) Positive and negative effects of widespread badger culling on tuberculosis in cattle. *Nature* 439(7078), 843-846.

Drexler J. F., Corman V. M. & Drosten C. (2014) Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antiviral Research* 101(0), 45-56.

Drosten C. (2013) Virus ecology: a gap between detection and prediction. *Emerging Microbes & Infections* 2(5), e31.

Drosten C., Günther S., Preiser W., van der Werf S., Brodt H. R., Becker S., Rabenau H., Panning M., Kolesnikova L., Fouchier R. A., Berger A., Burguière A. M., Cinatl J., Eickmann M., Escriou N., Grywna K., Kramme S., Manuguerra J. C., Müller S., Rickerts V., Stürmer M., Vieth S., Klenk H. D., Osterhaus A. D., Schmitz H. & Doerr H. W. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New England Journal of Medicine* 348(20), 1967-1976.

Fenton M. B. & Simmons N. B. (2015) *Bats: A World of Science and Mystery*. University of Chicago Press.

Jones G. & Barlow K. (2004) Cryptic species of echolocating bats. In: *Echolocation in bats and dolphins* (eds J. A. Thomas, C. F. Moss and M. Vater) pp. 345-349. University of Chicago Press, London.

Jones K., Patel N., Levy M., Storeygard A., Balk D., Gittleman J. & Daszak P. (2008) Global trends in emerging infectious diseases. *Nature* 451(7181), 990-993.

Keesing F., Belden L., Daszak P., Dobson A., Harvell C., Holt R., Hudson P., Jolles A., Jones K., Mitchell C., Myers S., Bogich T. & Ostfeld R. (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468(7324), 647-652.

Kunz T. H., Braun de Torrez E., Bauer D., Lobova T. & Fleming T. H. (2011) Ecosystem services provided by bats. *Annals of the New York Academy of Sciences* 1223(1), 1-38.

Leendertz S. A. J., Gogarten J. F., Düx A., Calvignac-Spencer S. & Leendertz F. H. (2015) Assessing the evidence supporting fruit bats as the primary reservoirs for Ebola viruses. *EcoHealth*, 1-8.

Leroy E. M., Epelboin A., Mondonge V., Pourrut X., Gonzalez J.-P., Muyembe-Tamfum J.-J. & Formenty P. (2009) Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector-Borne and Zoonotic Diseases* 9(6), 723-728.

Li W., Shi Z., Yu M., Ren W., Smith C., Epstein J., Wang H., Crameri G., Hu Z., Zhang H., Zhang J., McEachern J., Field H., Daszak P., Eaton B. T., Zhang S. & Wang L. F. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310(5748), 676-679.

Luis A. D., Hayman D. T. S., O'Shea T. J., Cryan P. M., Gilbert A. T., Pulliam J. R. C., Mills J. N., Timonin M. E., Willis C. K. R., Cunningham A. A., Fooks A. R., Rupprecht C. E., Wood J. L. N. & Webb C. T. (2013) A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proceedings of the Royal Society of London B Biological Sciences* 280(1756), 20122753.

McDonald R. A., Delahay R. J., Carter S. P., Smith G. C. & Cheeseman C. L. (2008) Perturbing implications of wildlife ecology for disease control. *Trends in Ecology & Evolution* 23(2), 53-56.

Mickleburgh S. P., Hutson A. M. & Racey P. A. (2002) A review of the global conservation status of bats. *Oryx* 36(01), 18-34.

Monadjem A., Taylor P. J., Cotterill W. & Schoeman M. (2010) *Bats of Southern and Central Africa: A Biogeographic and Taxonomic Synthesis*. Wits University Press, Johannesburg.

Morse S. S. (1995) Factors in the emergence of infectious diseases. *Emerging Infectious Diseases* 1(1), 7-15.

Patz J., Daszak P., Tabor G., Aguirre A., Pearl M., Epstein J., Wolfe N., Kilpatrick A., Foufopoulos J. & Molyneux D. (2004) Unhealthy landscapes: policy recommendations on land use change and infectious disease emergence. *Environmental Health Perspectives* 112(10), 1092-1098.

Pfefferle S., Oppong S., Drexler J., Gloza-Rausch F., Ipsen A., Seebens A., Müller M., Annan A., Vallo P., Adu-Sarkodie Y., Kruppa T. & Drosten C. (2009) Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerging Infectious Diseases* 15(9), 1377-1384.

Plowright R. K., Eby P., Hudson P. J., Smith I. L., Westcott D., Bryden W. L., Middleton D., Reid P. A., McFarlane R. A. & Martin G. (2015) Ecological dynamics of emerging bat virus spillover. *Proceedings of the Royal Society B: Biological Sciences* 282(1798), 2014-2124.

Saéz A. M., Weiss S., Nowak K., Lapeyre V., Zimmermann F., Düx A., Kühl H. S., Kaba M., Regnaut S. & Merkel K. (2015) Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Molecular Medicine* 7(1), 17-23.

Saif L. (2004) Animal coronaviruses: what can they teach us about the severe acute respiratory syndrome? *Revue scientifique et technique (International Office of Epizootics)* 23(2), 643-660.

Smith K. F., Acevedo-Whitehouse K. & Pedersen A. B. (2009) The role of infectious diseases in biological conservation. *Animal Conservation* 12(1), 1-12.

Streicker D. G., Recuenco S., Valderrama W., Gomez Benavides J., Vargas I., Pacheco V., Condori Condori R. E., Montgomery J., Rupprecht C. E., Rohani P. & Altizer S. (2012) Ecological and anthropogenic drivers of rabies exposure in vampire bats: implications for

transmission and control. *Proceedings of the Royal Society of London B Biological Sciences* 279(1742), 3384-3392.

Swanepoel R., Smit S., Rollin P., Formenty P., Leman P., Kemp A., Burt F., Grobbelaar A., Croft J. & Bausch D. (2007) Studies of reservoir hosts for Marburg virus. *Emerging Infectious Diseases* 13(12), 1847-1851.

Tong S., Conrardy C., Ruone S., Kuzmin I. V., Guo X., Tao Y., Niezgoda M., Haynes L., Agwanda B. & Breiman R. F. (2009) Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerging Infectious Diseases* 15(3), 482-485.

Vijaykrishna D., Smith G., Zhang J., Peiris J., Chen H. & Guan Y. (2007) Evolutionary insights into the ecology of coronaviruses. *Journal of Virology* 81(8), 4012-4020.

Woo P. C. Y., Lau S. K. P., Huang Y. & Yuen K.-Y. (2009) Coronavirus diversity, phylogeny and interspecies jumping. *Experimental Biology and Medicine* 234(10), 1117-1127.

Woo P. C. Y., Wang M., Lau S. K. P., Xu H., Poon R. W. S., Guo R., Wong B. H. L., Gao K., Tsoi H.-W., Huang Y., Li K. S. M., Lam C. S. F., Chan K.-H., Zheng B.-J. & Yuen K.-Y. (2007) Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *Journal of Virology* 81(4), 1574-1585.

Woodroffe R., Donnelly C. A., Cox D., Bourne F., Cheeseman C., Delahay R., Gettinby G., Mcinerney J. P. & Morrison W. (2006) Effects of culling on badger *Meles meles* spatial organization: implications for the control of bovine tuberculosis. *Journal of Applied Ecology* 43(1), 1-10.

Zaki A. M., van Boheemen S., Bestebroer T. M., Osterhaus A. D. M. E. & Fouchier R. A. M. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine* 367, 1814-1820.

Zhang G., Cowled C., Shi Z., Huang Z., Bishop-Lilly K. A., Fang X., Wynne J. W., Xiong Z., Baker M. L., Zhao W., Tachedjian M., Zhu Y., Zhou P., Jiang X., Ng J., Yang L., Wu

L., Xiao J., Feng Y., Chen Y., Sun X., Zhang Y., Marsh G. A., Crameri G., Broder C. C., Frey K. G., Wang L.-F. & Wang J. (2013) Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* 339(6116), 456-460

CHAPTER 2
MERS-CoV-related viruses in bats in Ghana and Europe

A. A. Annan#, H. J. Baldwin#, V. M. Corman#, S. M. Klose, M. Owusu, E. E. Nkrumah, E. K. Badu, P. Anti, O. Agbenyega, B. Meyer, S. K. Oppong, Y. Adu-Sarkodie, E. K. V. Kalko, P. Lina, E. Godlevska, C. B. E. M. Reusken, A. Seebens, F. Gloza-Rauch, P. Vallo, M. Tschapka, C. Drosten and J. F. Drexler. (2013) *Emerging Infectious Diseases*, 19(3): 456-459

# these authors contributed equally

ABSTRACT

We screened faecal specimens of 4,758 bats from Ghana and 272 bats from four European countries for betacoronaviruses. Viruses related to the novel human betacoronavirus MERS-CoV (formerly known as hCoV-EMC/2012) were detected in 46 (24.9%) of 185 *Nycteris* spp and 40 (14.7%) of 272 *Pipistrellus* spp. Their genetic relatedness indicate that MERS-CoV originated from bats.

INTRODUCTION

Coronaviruses (CoVs) are enveloped viruses with a positive-sense, single-stranded RNA genome (Woo *et al.* 2009). CoVs are classified into four genera: *Alphacoronavirus*, *Betacoronavirus* (grouped further into clades 2a–2d), *Gammacoronavirus*, and *Deltacoronavirus*. Two human coronaviruses (hCoVs), termed hCoV-OC43 and -229E, have been known since the 1960s and cause chiefly mild respiratory disease (Saif 2004). In 2002–2003, an outbreak of severe acute respiratory syndrome (SARS) leading to ≈850 deaths was caused by a novel group 2b betacoronavirus, SARS-CoV (Drosten *et al.* 2003). A likely animal reservoir for SARS-CoV was identified in rhinolophid bats (Drexler *et al.* 2010; Li *et al.* 2005). In the aftermath of the SARS pandemic, two hCoVs, termed hCoV-NL63 and -HKU1, and numerous novel bat CoVs were described.

In September 2012, health authorities worldwide were notified of two cases of severe respiratory disease caused by a novel hCoV (Corman *et al.* 2012; Zaki *et al.* 2012). Originally termed hCoV EMC/2012, this virus is now known as Middle East respiratory syndrome coronavirus (MERS-CoV). Since its emergence, there have been major outbreaks in Saudi Arabia and the Republic of Korea, with more than 1200 cases and 500 deaths worldwide (>40%; WHO 2015). MERS-CoV was found to be related to the 2c betacoronavirus clade, which had only been known to contain *Tylonycteris* bat coronavirus HKU4 and *Pipistrellus* bat coronavirus HKU5 (Woo *et al.* 2006).

We previously identified highly diversified alphacoronaviruses and betacoronaviruses, but not clade 2c betacoronaviruses, in bats from Ghana (Pfefferle *et al.* 2009). We also identified sequence fragments from a 2c betacoronavirus from one *Pipistrellus pipistrellus* in Europe (Reusken *et al.* 2010). In this study, we analyzed an extended sample of 4,758 bats from Ghana and 272 bats from four European countries.

THE STUDY

Faecal specimens were collected from ten bat species in Ghana and four *Pipistrellus* species in Europe (Table 1). Bats were caught during 2009–2011 with mist nets, as described (Pfefferle *et al.* 2009), in seven locations across Ghana and five areas in Germany, the Netherlands, Romania, and Ukraine (Figure 1). The species, age, sex, reproductive status, and morphologic measurements of the bats were recorded. Faecal pellets were collected and suspended in RNAlater Stabilization Reagent (QIAGEN, Hilden, Germany). RNA was purified as described (Drexler *et al.* 2012). CoV was detected by using nested reverse transcription PCR (RT-PCR) targeting the *RNA-dependent RNA polymerase* (*RdRp*) gene (de Souza Luna *et al.* 2007) (see Table 1 for assay oligonucleotides).

A novel CoV was detected in insectivorous *Nycteris* cf. *gambiensis* specimens (Technical Appendix; GenBank accession nos. JX899382–JX899384). A real-time RT- PCR was designed to permit sensitive and quantitative detection of this CoV (Table 1). Only *N.* cf. *gambiensis* bats were positive for CoV (46 [24.9%] of 185 specimens) (Table 1). Demographic factors predictive of CoV in captured *N.* cf. *gambiensis* were assessed. Juvenile bats and lactating females were significantly more likely to be CoV-infected than were adult and non-lactating female bats, respectively (Table 2). Virus concentrations in faeces from *N.* cf. *gambiensis* were high (median 412,951 RNA copies/g range 323–150,000,000 copies/g).

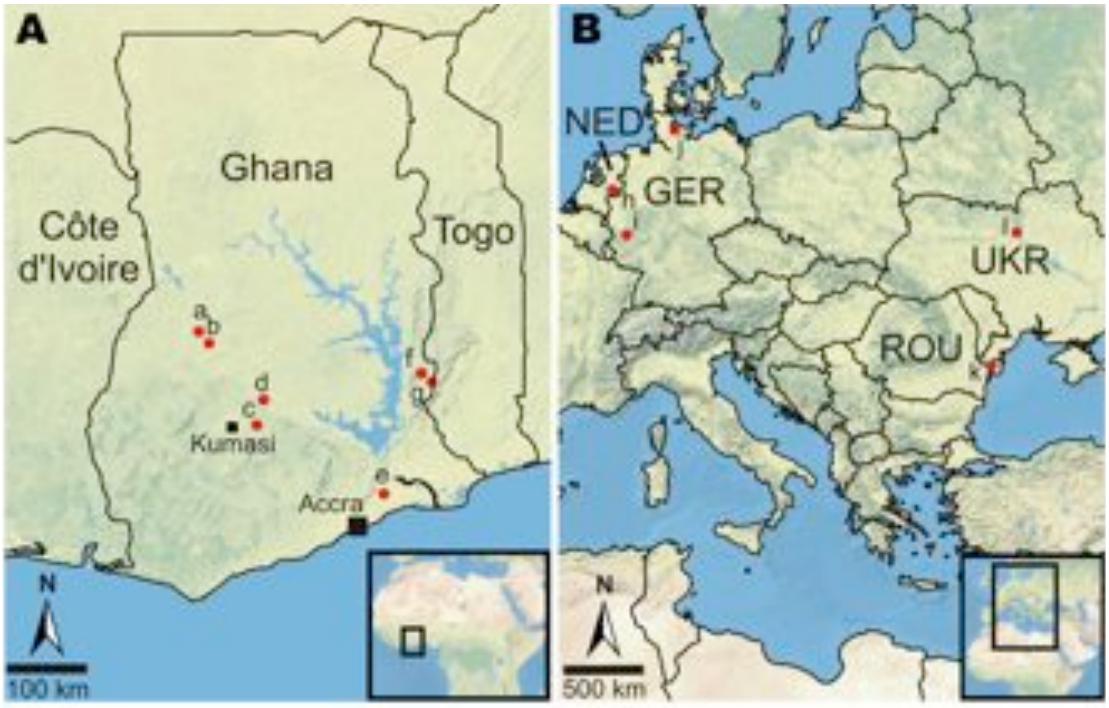


Figure 1. Location of bat sampling sites in Ghana and Europe. The 7 sites in Ghana (A) and the 5 areas in Europe (B) are marked with dots and numbered from west to east. a) Buoyem (N7°43′24.899″ W1°59′16.501″); b) Forikrom (N7°35′23.1″ W1°52′30.299″); c) Bobiri (N6°41′13.56″ W1°20′38.94″); d) Kwamang (N6°58′0.001″ W1°16′ 0.001″); e) Shai Hills (N5°55′44.4″ E0°4′30″); f) Akpafu Todzi (N7°15′43.099″ E0°29′29.501″); g) Likpe Todome (N7°9′50.198″ E0°36′28.501″); h) Province Gelderland, NED (N52°1′46.859″ E6°13′4.908″); i) Eifel, Rhineland-Palatinate, GER (N50°20′5.316″ E7°14′30.912″); j) Holstein, Schleswig-Holstein, GER (N54°14′51.271″ E10°4′3.347″); k) Tulcea county, ROU (N45°12′0.00″ E29°0′0.00″); l) Kiev region, UKR (N50°27′0.324″ E30°31′24.24″). NED, the Netherlands; GER, Germany; ROU, Romania; UKR, Ukraine.

The 398-bp CoV *RdRp* screening fragment was extended to 816 bp, as described (Drexler *et al.* 2010), to enable more reliable taxonomic classification. We previously established *RdRp*-grouping units (RGU) as a taxonomic surrogate to enable prediction of CoV species on the basis of this 816-bp fragment when no full genome sequences could be obtained. According to our classification, the amino acid sequences in the translated 816-bp fragment of the tentative betacoronavirus species (RGU) differed from each other by at least 6.3% (Drexler *et al.* 2010). The new *Nycteris* bat CoV differed from the 2c-prototype viruses HKU4 and HKU5 by 8.8%–9.6% and from MERS-CoV by 7.5% and thus constituted a novel RGU. A partial *RdRp*

sequence fragment of a *P. pipistrellus* bat CoV from the Netherlands, termed VM314 (described in Reusken *et al.* (2010)), was completed toward the 816-bp fragment to refine the RGU classification of MERS-CoV. MERS-CoV differed from VM314 by only 1.8%.

Table 1. Overview of bats tested for 2c-betacoronaviruses in Ghana

Area **Species**	***n* (positives)** **[%]**	**Juv/Ad†**	**F/M‡**	**Location§ (*n*/positives)**
Ghana				
Coleura afra	108 (0)	2/105	46/59	a,b,e
Hipposideros abae	604 (0)	55/548	207/341	a,b,d,f
H. cf. *gigas*	28 (0)	7/19	8/11	a,b,d
H. fuliginosus	1 (0)	1/0	-	c
H. jonesi	31 (0)	6/25	1/24	c,d
H. cf. *ruber*	3763 (0)	674/3078	1109/1969	a,b,c,d,f,g
Nycteris cf. *gambiensis*	185 (46) [24.9]	22/161¶	79/82	a# (5/2),b# (65/15),d# (104/29),f(1/0)
Rhinolophus alcyone	4 (0)	2/2	1/1	c
R. landeri	13 (0)	3/10	2/8	b,d,f
Taphozous perforatus	21 (0)	3/18	0/18	e
Total	4758 (46) [1.0]			
Europe	7 (0)			
Pipistrellus kuhlii	82 (30) [36.6]	Unknown	3/3	l
P. nathusii	42 (1) [2.4]	15/65	38/43	j (2/0), k# (74/29), l# (6/1)
P. pipistrellus	141 (9) [6.4]	17/25	19/21	i (29/0), k# (7/1), h (6/0)
P. pygmaeus	272 (40) [14.7]	11/127	83/55	j (44/0), k# (91/9), l (6/0)
Total	108 (0)			

The real-time reverse transcription PCR (Ghana) used oligonucleotides 2c-rtF, 5′-GCACTGTTGCTGGTGTCTCTATTCT-3′, 2crtR, 5′- GCCTCTAGTGGCAGCCATACTT-3′ and 2c-rtP, JOE-TGACAAATCGCCAATACCATCAAAAGATGC-BHQ1 and the Pan2c-heminested assay (Europe) used oligonucleotides Pan2cRdRP-R, 5′-GCATWGCNCWGTCACACTTAGG-3′; Pan2cRdRP-Rnest, 5′-CACTTAGGRTARTCCCAWCCCA-3′; and Pan2cRdRp-FWD, 5′-TGCTATWAGTGCTAAGAATAGRGC-3′. †Excludes bats (all coronavirus-negative) that were missing data for age. ‡ Excludes bats that were missing data for sex. §a, Buoyem; b, Forikrom; c, Bobiri; d, Kwamang; e, Shai Hills; f, Akpafu Todzi, g, Likpe Todome; h, Province Gelderland; i, Eifel area; j, Holstein area; k, Tulcea county; l, Kiev region; GPS coordinates are shown in Figure 1. ¶For two animals, no data on age were available. #Locations in which coronavirus 2c–positive bats were found.

Because of the genetic similarity between MERS-CoV and VM314, we specifically investigated *Pipistrellus* spp. from four European countries for 2c betacoronaviruses. We detected betacoronaviruses in 40 (14.7%) of 272 *P. pipistrellus, P. nathusii, and P. pygmaeus* from the

Netherlands, Romania, and Ukraine (Table 1; GenBank accession nos. KC243390-KC243392) that were closely related to VM314. The VM314-associated *Pipistrellus* bat betacoronaviruses differed from MERS-CoV by 1.8%. The difference between MERS-CoV and HKU5 was 5.5% – 5.9%. In summary, HKU5, MERS-CoV, and the VM314- associated clade form one RGU according to our classification system, and the VM314-*Pipistrellus* bat clade contains the closest relatives of MERS-CoV. HKU4 and the *Nycteris* CoV define two separate tentative species in close equidistant relationship.

Table 2. Possible factors predictive of 2c betacoronavirus detection in *N.*cf. *gambiensis*

Factor	Category	*n*	No. of CoV positives (%)	χ^2	*p*-value	OR	OR (95% CI)	
							Min	Max
Age	Juvenile	22	10 (45.4)	5.49	0.02	2.89	1.16	7.24
	Adult	161	36 (22.4)					
Sex	Female	79	16 (20.3)	0.01	0.91	1.04	0.50	2.17
	Male	82	20 (24.4)					
Lactation (♀)	Lactating	25	11 (44.0)	12.77	0.0004	7.70	2.29	25.89
	Non-lactating	54	5 (9.3)					
Gravidity (♀)	Gravid	13	0 (0.00)	3.95	0.06[§]	0.00	-	-
	Non-gravid	66	16 (24.2)					
Reproductive status (♂)	Active	56	15 (26.8)	0.55	0.46	1.54	0.49	4.81
	Non-reproductive	26	5 (19.2)					

All analyses were done using uncorrected chi2 tests (two-tailed) in EpiInfo V7 (http://wwwn.cdc.gov/epiinfo/7/index.htm) with the exception of the gravity parameter, because one of the expected values was less than 5. All analyses except "Age" excluded juveniles. [§]fisher exact test.

n = number of bats tested, OR = odds ratio

We conducted a Bayesian phylogenetic analysis. In this analysis, the *Nycteris* bat CoV clustered as a phylogenetically basal sister clade with HKU4, HKU5, and MERS-CoV and the associated European *Pipistrellus* viruses (Figure 2A).

To confirm the *RdRp*-based classification, we amplified the complete glycoprotein-encoding *Spike* gene and sequenced it for the novel *Nycteris* bat virus. The phylogenetically basal position of the novel *Nycteris* bat virus within the 2c clade resembled that in the CoV *RdRp* gene (Figure 2B). Partial sequences that could be obtained from the 3′-end of the *Spike* gene of three

2c *Pipistrellus* bat betacoronaviruses confirmed their relatedness to MERS-CoV (Figure 2C, MERS-CoV represented as EMC/2012).

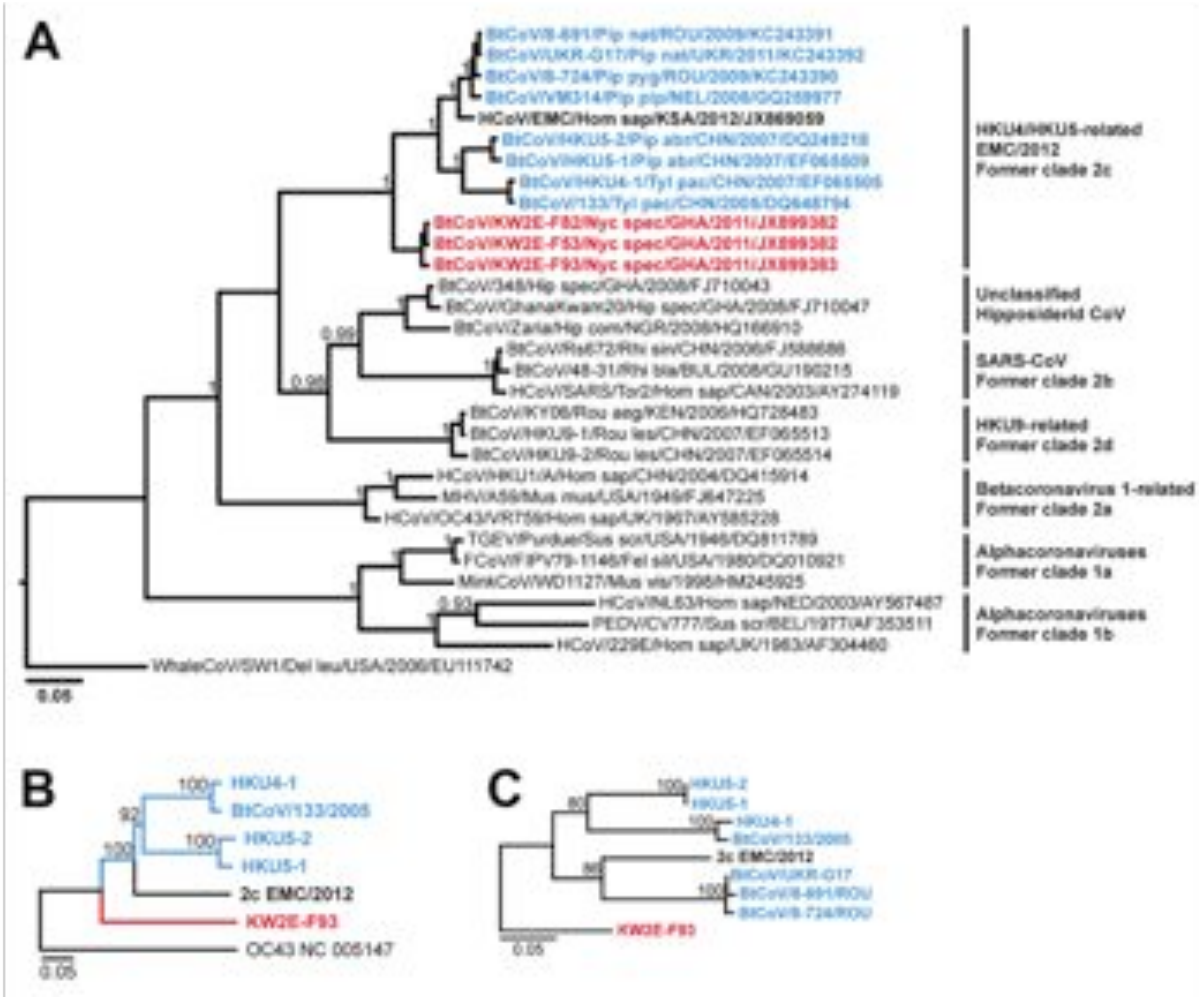


Figure 2. *RNA-dependent RNA polymerase* (*RdRp*) gene and *Spike* gene phylogenies including the novel betacoronaviruses from bats in Ghana and Europe. A) Bayesian phylogeny of an 816-nt *RdRp* gene sequence fragment corresponding to positions 14781–15596 in severe acute respiratory syndrome coronavirus (SARS-CoV) strain Frankfurt 1 (GenBank accession no. AY291315). Data were analyzed with MRBAYES version 3.1 (http://mrbayes.sourceforge.net/) by using a WAG amino acid substitution model and 4 million generations sampled every 100 steps. Trees were annotated by using a burn-in of 10,000 and visualized with FIGTREE version 1.6.1 from the BEAST package (www.beast.bio.ed.ac.uk). A whale gammacoronavirus was used as an outgroup. The novel *Nycteris* bat viruses are shown in boldface and red, the novel *Pipistrellus* bat viruses and other bat CoVs in the 2c clade are shown in boldface and cyan, and the novel human betacoronavirus EMC/2012 (MERS-CoV) is shown in boldface. Values at deep nodes represent statistical support of grouping by posterior probabilities. CoV clades are depicted to the right of taxa. B) Phylogeny of the complete *Spike* gene of clade 2c CoVs determined by using the neighbour-joining method with an amino acid percentage distance substitution model and the complete deletion

option in MEGA5 (www.megasoftware.net). The Nycteris CoV *Spike* gene was equidistant from other 2c-CoV *Spike* genes with 45.6%–46.8% aa divergence. Human coronavirus (hCoV)–OC43 was used as an outgroup. No complete *Spike* gene sequence was available for VM314 or the novel *Pipistrellus* bat CoVs. Scale bar represents percentage amino acid distance. The analysis comprised 1,731 aa residues. C) Phylogeny of the partial *Spike* gene of clade 2c CoVs, including the novel CoVs of *Pipistrellus* bats from Europe, determined by using a nucleotide distance substitution model and the complete deletion option in MEGA5. Scale bar represents percentage nucleotide distance. The analysis comprised 131 nt corresponding to positions 25378–25517 in MERS-CoV. Oligonucleotide sequences of primers used to amplify full and partial *Spike* gene sequences are available on request from the authors. Values at deep nodes in B and C represent statistical support of grouping by percentage of 1,000 bootstrap replicates. GenBank accession numbers for the complete and partial *Spike* genes correspond to those given in panel A for the *RdRp* gene.

CONCLUSIONS

We detected novel clade 2c betacoronaviruses in *Nycteris* cf. *gambiensis* in Ghana and *Pipistrellus* spp. in Europe that are phylogenetically related to the novel MERS-CoV. All previously known 2c bat CoVs originated from vespertilionid bats: VM314 originated from *P. pipistrellus* from the Netherlands and HKU4 and HKU5 originated from *Tylonycteris pachypus* and *P. abramus*, respectively, from the People's Republic of China. The *Nycteris* bat virus in Africa extends this bat CoV clade over two different host families, Nycteridae and Vespertilionidae (Technical Appendix). Detection of genetically related betacoronaviruses in bats from Africa and Eurasia parallels detection of SARS-CoV in rhinolophid bats from Eurasia and related betacoronaviruses in hipposiderid bats from Africa (Pfefferle *et al.* 2009).

The relatedness of MERS-CoV to CoVs hosted by *Pipistrellus* bats at high prevalence across different European countries and the occurrence of HKU5 in bats of this genus from China highlight the possibility that *Pipistrellus* bats might indeed host close relatives of MERS-CoV. This suspicion is supported by observations that tentative bat CoV species (RGUs) are commonly detected within one host genus (Drexler *et al.* 2010). Within the Arabian Peninsula, the International Union for Conservation of Nature (www.iucn.org) lists 50 bat species,

including the pipistrelloids *P . arabicus*, *P . ariel*, *P. kuhlii*, *P . pipistrellus*, *P. rueppellii*, and *Hypsugo savii*. Because of the epidemiologic link of EMC/2012 with the Arabian Peninsula (Corman *et al.* 2012; Zaki *et al.* 2012), bats from this area should be specifically screened.

The genomic data suggest that MERS-CoV, like hCoV- 229E and SARS-CoV, might be another human CoV for which an animal reservoir of closely related viruses could exist in Old World insectivorous bats (Li *et al.* 2005; Pfefferle *et al.* 2009). Whether cross-order (e.g., chiropteran, carnivore, primate) host switches, such as suspected for SARS-CoV, have occurred for 2c clade bat CoVs remains unknown. However, we showed previously that CoVs are massively amplified in bat maternity colonies in temperate climates (Drexler *et al.* 2011). This amplification also might apply to the *Nycteris* bat CoV because, as shown previously for vespertilionid bats from temperate climates (Gloza-Rausch *et al.* 2008), detection rates of CoV are significantly higher among juvenile and lactating *N.* cf. *gambiensis*. In light of the observed high virus concentrations, the use of water from bat caves and bat guano as fertilizer for farming and the hunting of bats as wild game throughout Africa (Mickleburgh *et al.* 2009) may facilitate host switching events. To our knowledge, no CoV has been isolated directly from bats. Further studies should still include isolation attempts to obtain full virus genomes and to identify virulence factors that may contribute to the high pathogenicity of MERS-CoV (Zaki *et al.* 2012).

ACKNOWLEDGEMENTS

We thank Sebastian Brünink, Tobias Bleicker, and Monika Eschbach-Bludau for technical assistance. We are grateful to Ioan Coroiu, Carsten Dense, Regina Klüppel-Hellmann, Anda Culisier, Danny Culisier, Sabrina Stölting, the volunteers at the Bonn Consortium for Bat Conservation, Andreas Kiefer, Manfred Braun, Isaac Mawusi Adanyeguh, Lucinda Kirkpatrick, Mac Elikem Nutsuakor, David Ofori Agyei, Sarah Koschnicke, Julia Morrison, Emmanual Asare, and Thomas Kruppa for their help during the organization and conduct of field work. We thank Anna Marie Corman for assistance with geographic information processing. For all capturing,

sampling, and exportation of bat specimens, we obtained permission from the respective countries' authorities. This study was supported by the European Union FP7 projects EMPERIE (contract number 223498) and ANTIGONE (contract number 278976) and by the German Research Foundation (DFG grant DR 772/3-1, KA1241/18-1).

REFERENCES

Corman V. M., Eckerle I., Bleicker T., Zaki A., Landt O., Eschbach-Bludau M., van Boheemen S., Gopal R., Ballhause M., Bestebroer T. M., Muth D., Muller M. A., Drexler J. F., Zambon M., Osterhaus A. D., Fouchier R. M. & Drosten C. (2012) Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction. *Eurosurveillance* 17(39), 30-35.

de Souza Luna L., Heiser V., Regamey N., Panning M., Drexler J., Mulangu S., Poon L., Baumgarte S., Haijema B. & Kaiser L. (2007) Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *Journal of Clinical Microbiology* 45(3), 1049-1052.

Drexler J. F., Corman V. M., Muller M. A., Maganga G. D., Vallo P., Binger T., Gloza-Rausch F., Rasche A., Yordanov S., Seebens A., Oppong S., Sarkodie Y. A., Pongombo C., Lukashev A. N., Schmidt-Chanasit J., Stocker A., Carneiro A. J. B., Erbar S., Maisner A., Fronhoffs F., Buettner R., Kalko E. K. V., Kruppa T., Franke C. R., Kallies R., Yandoko E. R. N., Herrler G., Reusken C., Hassanin A., Kruger D. H., Matthee S., Ulrich R. G., Leroy E. M. & Drosten C. (2012) Bats host major mammalian paramyxoviruses. *Nature Communications* 3, 796.

Drexler J. F., Corman V. M., Wegner T., Tateno A. F., Zerbinati R. M., Gloza-Rausch F., Seebens A., Müller M. A. & Drosten C. (2011) Amplification of emerging viruses in a bat colony. *Emerging Infectious Diseases* 17(3), 449-456.

Drexler J. F., Gloza-Rausch F., Glende J., Corman V. M., Muth D., Goettsche M., Seebens A., Niedrig M., Pfefferle S. & Yordanov S. (2010) Genomic characterization of SARS-related coronavirus in European bats and classification of coronaviruses based on partial *RNA-dependent RNA polymerase* gene sequences. *Journal of Virology* 84(21), 11336-11349.

Drosten C., Günther S., Preiser W., van der Werf S., Brodt H. R., Becker S., Rabenau H., Panning M., Kolesnikova L., Fouchier R. A., Berger A., Burguière A. M., Cinatl J., Eickmann M., Escriou N., Grywna K., Kramme S., Manuguerra J. C., Müller S., Rickerts V., Stürmer M., Vieth S., Klenk H. D., Osterhaus A. D., Schmitz H. & Doerr H. W. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New England Journal of Medicine* 348(20), 1967-1976.

Gloza-Rausch F., Ipsen A., Seebens A., Göttsche M., Panning M., Drexler J., Petersen N., Annan A., Grywna K. & Müller M. (2008) Detection and prevalence patterns of group I coronaviruses in bats, northern Germany. *Emerging Infectious Diseases* 14(4), 626-631.

Li W., Shi Z., Yu M., Ren W., Smith C., Epstein J., Wang H., Crameri G., Hu Z., Zhang H., Zhang J., McEachern J., Field H., Daszak P., Eaton B. T., Zhang S. & Wang L. F. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310(5748), 676-679.

Mickleburgh S., Waylen K. & Racey P. (2009) Bats as bushmeat: a global review. *Oryx* 43(2), 217-234.

Pfefferle S., Oppong S., Drexler J., Gloza-Rausch F., Ipsen A., Seebens A., Müller M., Annan A., Vallo P., Adu-Sarkodie Y., Kruppa T. & Drosten C. (2009) Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerging Infectious Diseases* 15(9), 1377-1384.

Reusken C. B. E. M., Lina P. H. C., Pielaat A., de Vries A., Dam-Deisz C., Adema J., Drexler J. F., Drosten C. & Kooi E. A. (2010) Circulation of group 2 coronaviruses in a bat species

common to urban areas in Western Europe. *Vector-Borne and Zoonotic Diseases* 10(8), 785-791.

Saif L. (2004) Animal coronaviruses: what can they teach us about the severe acute respiratory syndrome? *Revue scientifique et technique (International Office of Epizootics)* 23(2), 643-660.

Woo P., Lau S., Li K., Poon R., Wong B., Tsoi H., Yip B., Huang Y., Chan K. & Yuen K. (2006) Molecular diversity of coronaviruses in bats. *Virology* 351(1), 180-187.

Woo P. C. Y., Lau S. K. P., Huang Y. & Yuen K.-Y. (2009) Coronavirus diversity, phylogeny and interspecies jumping. *Experimental Biology and Medicine* 234(10), 1117-1127.

World Health Organisation. (2015) Middle East respiratory syndrome coronavirus (MERS-CoV) – Saudi Arabia. *Disease outbreak news*, 13 November 2015.

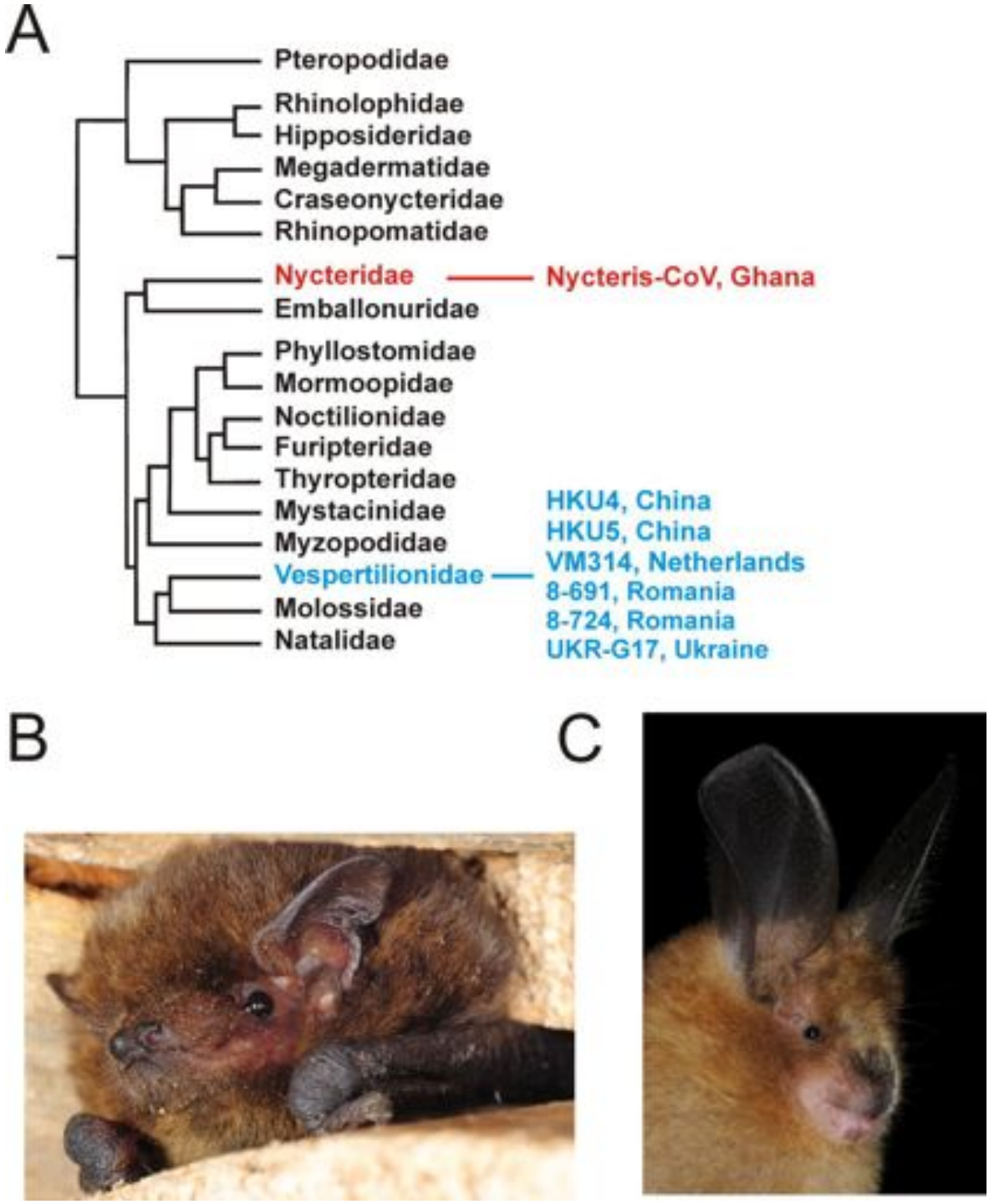
Zaki A. M., van Boheemen S., Bestebroer T. M., Osterhaus A. D. M. E. & Fouchier R. A. M. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine* 367, 1814-1820.

TECHNICAL APPENDIX

Technical appendix figure 1. (opposite) Bat evolutionary lineages and species in which novel group 2c betacoronaviruses were detected, Ghana and Europe. A) Bat evolutionary lineages in which novel group 2c betacoronaviruses were detected. Phylogeny adapted from Simmons *et al.* (2005). Bat families Nycteridae and Vespertilionidae and coronaviruses hosted by bats of these families detected in this and previous studies are shown in red and cyan, respectively. CoV, coronavirus. B) European *Pipistrellus nathusii*; photo by Florian Gloza-Rausch. C) Ghanaian *Nycteris* cf. *gambiensis*; photo by Marco Tschapka.

REFERENCES

Simmons *et al.* 2005. An Eocene big bang for bats. *Science* 307, 527-8.



Technical appendix figure 1. See description opposite.

CHAPTER 3
Widespread coronavirus infection in Ghanaian bats: seasonal, demographic and ecological factors influence infection risk

Image: Marco Tschapka

H. J. Baldwin, V. M. Corman, E. E. Nkrumah, E. K. Badu, P. Anti, A. A. Annan, M. Owusu, O. Agbenyega, S. K. Oppong, S. M. Klose, Y. Adu-Sarkodie, A. J. Stow, P. Vallo, E. K. V. Kalko, J. F. Drexler, C. Drosten and M. Tschapka.

Under review in *Ecography*

ABSTRACT

Bats are implicated in the emergence of several zoonotic diseases, including the coronaviruses (CoV) responsible for severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). Despite the considerable public health and economic cost of zoonotic CoV outbreaks, little research has examined the ecology of CoVs in bat populations. To examine individual and population-level risk factors for CoV infection in bats, we conducted a longitudinal study of CoV in cave dwelling bats in Ghana, West Africa. We report widespread CoV infection in African bats, with six of 17 bat species infected with CoVs belonging to several *Alphacoronavirus* and *Betacoronavirus* lineages. Juvenile bats had substantially higher risk of infection than adults. There was a strong temporal association, with April-July having generally higher detection rates, which may be compatible with an important role of juveniles and seasonal roosting behaviour for CoV amplification. CoV co-infection and ectoparasitic infection showed varying degrees of positive association with CoV infection, and there was evidence that lower body condition may increase risk. These findings provide new insights into the seasonal, demographic, ecological and processes that influence CoV infection dynamics in bats. We suggest avenues on which to focus future strategies for the prediction and prevention of zoonotic CoV outbreak, including avoiding consumption of juvenile bats and avoiding consumption and seasonally avoiding direct and indirect contact with bats.

Keywords: zoonosis, bats, Africa, coronavirus, ecology, epidemiology

INTRODUCTION

Emerging infectious diseases pose a threat to public health, animal agriculture and wildlife conservation (Daszak *et al.* 2000). The majority of emerging infectious diseases in humans are zoonoses, and two thirds of these originate in wildlife (Jones *et al.* 2008). Bats (Order Chiroptera) are reservoir hosts for many highly pathogenic zoonotic viruses, including Ebola, Marburg,

Hendra, Nipah, and the SARS coronavirus (Calisher *et al.* 2006). This propensity to host zoonotic viruses may be linked to chiropteran adaptations including high levels of interspecific roosting sympatry, sociality and gregariousness, longevity, high mobility due to flight, as well as genetic and functional differences in the immune system compared to other mammals (Calisher *et al.* 2006; Dobson 2005; Luis *et al.* 2013; Zhang *et al.* 2013).

As the extent of the diversity of bat-hosted viruses is revealed, bats have been found to be important animal reservoirs for coronaviruses (Drexler *et al.* 2014). Coronaviruses (CoV) are RNA viruses (Order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*) that affect a wide range of vertebrates, primarily infecting the respiratory and gastroenteric systems (Weiss and Navas-Martin 2005). The genera *Alphacoronavirus* and *Betacoronavirus* are associated with infection in mammals, whereas *Delta-* and *Gammacoronavirus* are found mainly in birds (Woo *et al.* 2009). Livestock CoVs such as transmissible gastroenteritis virus of swine, porcine epidemic diarrhoea virus, the bovine CoV and the infectious bronchitis virus are of significant relevance for the livestock industry (Saif 2004).

The six coronaviruses known to infect humans belong to the genera *Alphacoronavirus* (HCoV-NL63 and HCoV-229E) and *Betacoronavirus* (SARS-CoV, MERS-CoV, HCoV-OC43, HCoV-HKU1). Two of these human CoVs can be highly pathogenic. SARS-CoV was responsible for an outbreak of severe acute respiratory disease in 2002 and 2003 with up to 10% case fatality (WHO, 2004). MERS-CoV is the cause for an ongoing epidemic of severe respiratory disease associated with the Arabian peninsula (Zaki *et al.* 2012). Both SARS-CoV and MERS-CoV likely share a putative evolutionary origin in bats, including other animal intermediate hosts (Haagmans *et al.* 2014; Ithete *et al.* 2013; Lau *et al.* 2005; Li *et al.* 2005).

In the aftermath of the SARS epidemic, efforts to identify wildlife that host CoVs have intensified and many new bat CoVs have been described. Several CoVs have been identified in African bats, including close relatives of MERS-CoV, SARS-CoV, and HCoV-229E (Annan *et al.* 2013; Geldenhuys *et al.* 2013; Ithete *et al.* 2013; Pfefferle *et al.* 2009; Quan *et al.* 2010; Tao *et al.*

2012; Tong *et al.* 2009). Anthropogenic environmental change such as urbanization and human encroachment into wildlife habitat are important drivers of disease emergence, and are accelerating throughout Africa (Bradley and Altizer 2006; Daszak *et al.* 2001). Practices such as bushmeat consumption and cave visitation for resource extraction have been directly linked to serious human viral outbreaks on the continent (Amman *et al.* 2012; Leroy *et al.* 2009; Swanepoel *et al.* 2007), possibly including the recent catastrophic Ebola epidemic in West Africa (Saéz *et al.* 2015). There is therefore a pressing need for greater insight into the infection dynamics and ecology of CoVs in wildlife reservoirs.

The considerable effort undertaken in recent years to describe novel reservoir-borne viruses will not translate to prediction and prevention of zoonotic disease outbreaks without an understanding of the ecological mechanisms that drive virus emergence (Drosten 2013). Longitudinal and quantitative studies of viral reservoir ecology are needed to elucidate and model the risk factors that contribute to epizootic and zoonotic disease outbreaks and dynamics. Failure to consider the role of these factors can lead to ineffective, or even counter-productive management strategies with negative outcomes both for public health and wildlife conservation (Streicker *et al.* 2012; Woodroffe *et al.* 2006). Bats provide important ecosystem services (Kunz *et al.* 2011), and management decisions should be informed by host ecology in order to prevent zoonotic outbreaks while ensuring the conservation of bats and their habitats.

With the aim of increasing our limited understanding of CoV dynamics, we conducted a longitudinal study to evaluate the role of ecological and demographic factors in CoV infection risk in bats. We tested more than 7,000 bat faecal samples from 17 species for CoVs, collected at ten bat colonies in Ghana regularly over two years. This approach allowed us to characterize the infection dynamics of CoVs at the individual and population level.

Here we examine a number of factors that may influence host susceptibility and pathogen dynamics, including age and reproductive status, sex, body condition index, time of year, and co-infection by ectoparasites and other CoVs. Previous work on CoVs showed higher detection rates

for juvenile and lactating female bats (Annan *et al.* 2013; Gloza-Rausch *et al.* 2008). There may be seasonal differences in virus infection risk, as demonstrated for CoVs (Drexler *et al.* 2011; Gloza-Rausch *et al.* 2008; Osborne *et al.* 2011) and other bat viruses (Amman *et al.* 2012; Plowright *et al.* 2008; Serra-Cobo *et al.* 2013; Wacharapluesadee *et al.* 2010). We predicted that infection would be more likely in young bats and lactating females. We further predicted that a higher body condition index would reduce infection risk, while ectoparasites and other CoVs may be associated with higher risk (Plowright *et al.* 2008; Telfer *et al.* 2010).

MATERIALS AND METHODS

Sample collection

Faecal samples from a total of 7404 bats were collected over two years (August 2010 – August 2012) in Ghana, West Africa (figure 1, table S1). Most of our sampling sites were day roosts located in caves, abandoned mines or buildings in close proximity to human settlements, with a few capture sites located on forested or farming land. Core sites were sampled bimonthly, with a few sites sampled twice per year, and the remaining sites sampled opportunistically (table S1).

Bats were captured using mist nets set at approximately one hour after dusk until dawn. Animals were held in individual cloth bags until demographic and morphometric parameters were recorded. Juveniles were recognized by incomplete fusion of the fourth epiphyseal joint (Brunet-Rossinni and Wilkinson 2009). Ectoparasites were identified to subclass Acari (mites), family Streblidae (streblid flies), and family Nycteribiidae (nycteribiid flies). Faecal samples were collected and stored in RNAlater Stabilization Reagent (QIAGEN, Hilden, Germany).

The bats assigned to the currently recognized species *H. ruber* comprise three distinct mitochondrial lineages in this region (Vallo *et al.* 2008). A subset of individuals were sequenced at the cytochrome *b* gene to confirm the presence of all three lineages in our study sites, therefore we refer to animals keyed to this species as *H.* cf. *ruber*.

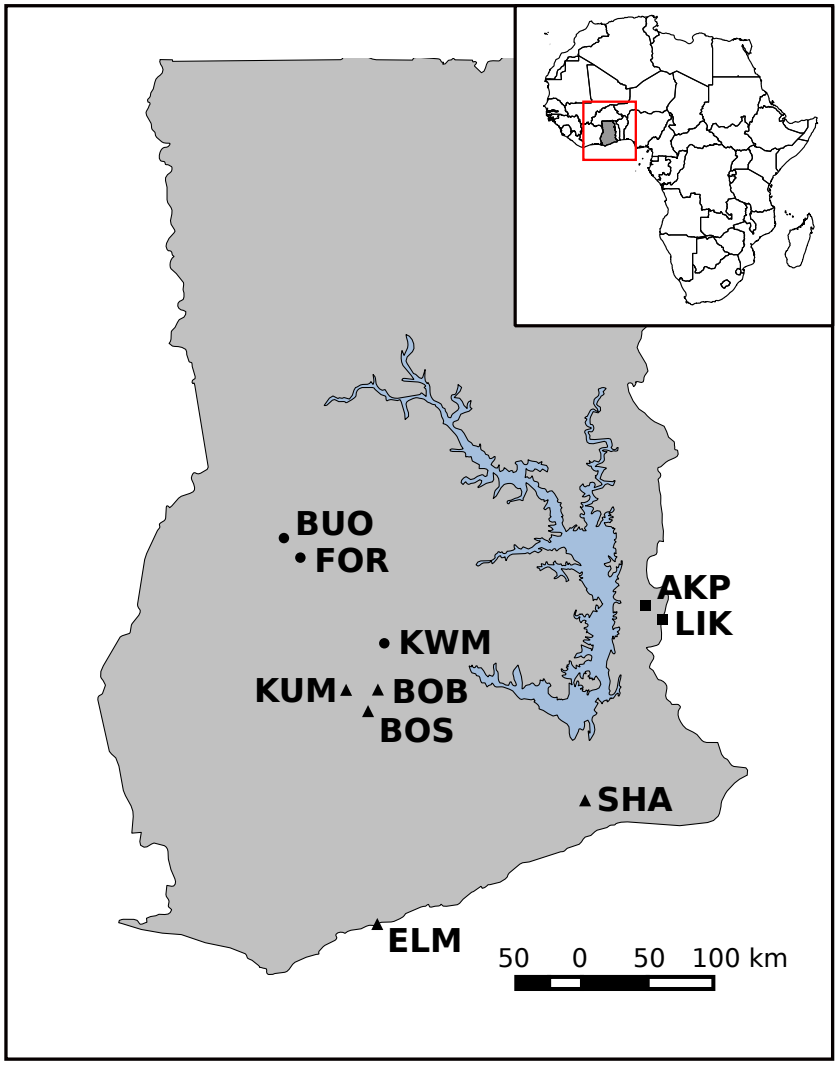


Figure 1. Sampling locations. Within some locations, samples were collected from multiple sites (see Table S1). BUO=Buoyem, FOR=Forikrom, KWM=Kwamang, AKP=Akpafu Todzi, LIK=Likpe Todome, KUM=Kumasi, BOB=Bobiri, BOS=Lake Bosumtwi, SHA=Shai Hills, ELM=Elmina. Circles, sampled bimonthly; squares, sampled semi-annually; triangles, sampled opportunistically

Viral RNA amplification

Faecal samples were analysed with reverse-transcriptase PCR (RT-PCR) after RNA purification as described elsewhere (Drexler *et al.* 2012). Five µl of RNA eluate were tested by lineage-specific real time reverse transcriptase (RT-) PCR assays designed to detect the diversity of four different CoV lineages described previously: Alpha229E, BetaBI, BetaBII and BetaC (also referred to as GhanaBt-CoVGrp1, GhanaBt-CoVGrp2 (BetaBI and BII), and 2c, respectively)(Annan *et al.*

2013; Corman *et al.* 2013b; Drexler *et al.* 2011; Drexler *et al.* 2010). Real-time RT-PCR protocols that were validated and performed in two multiplex reactions (table S2). In vitro transcribed RNA was generated for use as positive controls in each PCR run as described (Corman *et al.* 2013a).

Statistical analysis

In order to analyze the influence of ecological and demographic factors on CoV infections, data collected from three bat species (*H. abae, H.* cf. *ruber,* and *N.* cf. *gambiensis*) were used to fit generalized linear mixed models (GLMMs). Binomial GLMMs (logit link) were fitted to infection data from 2011 and 2012 using the lme4 package (Bates *et al.* 2013) in R (R Core Team 2013). Models were fitted for infection risk for each virus and bat species separately (see table S3 for global models.) Fixed effects for all models included age, sex, reproductive status, month captured, scaled body mass index (BMI), presence of ectoparasites, and co-infection with other coronaviruses. The scaled body mass index uses a standardised major axis regression method to avoid systematic bias towards larger individuals, and provides a more reliable indicator of true body condition than other mass indices (Peig and Green 2009). As age and sex are perfectly multicollinear with most values of reproductive status, global models alternatively included age and sex separately from reproductive status. Random effects for all models included site nested within location, year, and observer, to correct for any measurement bias. Records with missing values were excluded from the data sets. Pregnant females were excluded from the main analysis due the likelihood of confounding estimates for BMI, with parameter estimates for pregnant female calculated by fitting a second set of models from which BMI was excluded. Locations in which a virus was not present were excluded from the analysis of that virus. Variables that applied to less than 3% of cases were excluded from analyses.

Global models were tested for goodness-of-fit by calculating the concordance probability, or area under receiver operator curves (AUC) in the ROCR package in R (Sing *et al.* 2005).

Global models fits were tested against the null models using Akaike's adjusted Information Criterion (AICc; Burnham and Anderson 2002).

Multicollinearity and overdispersion were checked via adjusted generalized variance inflation factors (GVIF) from full models, excluding random effects, using the CAR package in R prior to analysis (Fox and Weisberg 2011). Adjusted GVIFs were below 2 for all models, and therefore collinearity was considered not to be an issue in this data set (Fox and Monette 1992).

An information-theoretic approach was used for model selection and multimodel inference using AICc (Burnham and Anderson 2002). Information-theoretic approaches, based on Kullback-Leibler (K-L) information, provide a quantitative measure of the strength of evidence for competing hypotheses by comparing and ranking competing models (Hegyi and Garamszegi 2011). The major advantage of I-T methods over null hypothesis testing is the ability to account for the uncertainty that is inherent in all statistical models.

Model selection and inference were based on all subsets of the global models and performed using the MuMIn package in R (Bartoń 2013). Models were ranked by their AICc value and Akaike weights, which estimates the likelihood of a model candidate representing the best model, taking into account both model fit and parsimony. Final parameter estimates of odds ratios were averaged over all models included in the top model set (ΔAICc < 2; Burnham and Anderson 2002). It should be noted that the ranking of models with correlated variables (reproductive status and sex) will potentially decrease model weights and increase the number of models in the top model set as they may overlap in explanatory power. Further, because of the high number of models in the model sets, the Δ2 cut-off may include spurious variables (Burnham and Anderson 2002). Therefore relative support for predictors was estimated by relative variable importance (RI), which is calculated as the sum of the Akaike weights across models including that variable, ranging from zero (no support) to one (maximum relative support) (Burnham and Anderson 2002). We consider variables with relative importance greater than 0.9 to have strong support, 0.7 to 0.9 moderate support, and weakly supported where

relative importance is greater than 0.4, with confidence intervals also used to gauge the strength of evidence. To ensure a balance of variables in the candidate sets, variable importances and model averages are calculated from separate model sets including either reproductive status or sex, based on the variable with highest relative importance. Differences in averaged values between these sets were negligible.

RESULTS

Of 7404 bats sampled from 17 species and five families, 1159 individuals from six species and two families were infected with four species of CoVs (1276 infections) (Table 1, Table S4). The Alpha229E, BetaBI and BetaBII-CoVs were found only in *Hipposideros* spp., and BetaC-CoV occurred exclusively in *N.* cf. *gambiensis*. Overall prevalence ranged from 0.8% (BetaC) to 8.7% (BetaBI), with the highest prevalence in a single bat species at 22.3% (BetaC, *N.* cf. *gambiensis*). Multiple infections occurred only in *Hipposideros* cf. *ruber* and *H. abae*, with double infections occurring in 1.7% of *H.* cf. *ruber* (19.2% of infections) and 0.9% *H. abae* (6.7% of infections), while triple infections occurred in five *H.* cf. *ruber* individuals (0.08% or 0.5% of infections). Juveniles had higher rates of multiple infection than adults, although this was significant only in *H. abae* (*H. abae*, 20.0% vs 2.7%, p=0.01; *H.* cf. *ruber*, 13.8% vs 9.5%, p=0.086; Pearson chi-square test, 10 000 MC replicates). Evidence of yearly variation was observed for BetaBII (increasing from 2010 to 2012) and BetaBI virus (decreasing 2011 to 2012).

In order to analyze the influence of ecological and demographic factors on CoV infection, data from the three bat species with highest CoV prevalence were used to fit generalized linear models. Confidence sets of the best-ranked regression models (ΔAICc < 2) can be found in table S5. The concordance probability for top models (AUC), a measure of model performance, ranged between 0.65 and 0.84 (table S5). Average estimates and relative variable importances of the parameters retained in confidence sets are presented in table 2. Parameter estimates for top fitted models are provided in tables S6 and S7.

Table 1. Overview of bats tested for coronaviruses

Species	Bat family	No. tested	No.[%] BetaC	No. [%] Alpha229E	No. [%] BetaBI	No. [%] BetaBII
Hipposideros abae	Hipposideridae	779	0	46 [5.9]	60 [7.7]	2[0.3]
H. fuliginosus	Hipposideridae	1	0	0	0	0
H. cf. *gigas*	Hipposideridae	49	0	0	1[2.0]	1[2.0]
H. jonesi	Hipposideridae	50	0	0	4[8.0]	1[2.0]
H. cf. *ruber*	Hipposideridae	5875	0	321[5.5]	585[10.0]	205[3.5]
H. caffer tephrus	Hipposideridae	128	0	6[4.7]	1[0.8]	0
Nycteris cf. *gambiensis*	Nycteridae	265	59[22.3]	0	0	0
Rhinolophus alcyone	Rhinolophidae	5	0	0	0	0
R. landeri	Rhinolophidae	14	0	0	0	0
Coleura afra	Emballonuridae	174	0	0	0	0
Taphozous perforatus	Emballonuridae	21	0	0	0	0
Lissonycteris angolensis	Pteropodidae	20	0	0	0	0
Rousettus aegyptiacus	Pteropodidae	2	0	0	0	0
Epomops buettikoferi	Pteropodidae	7	0	0	0	0
Micropteropus pusillus	Pteropodidae	3	0	0	0	0
Myonycteris torquata	Pteropodidae	5	0	0	0	0
Nanonycteris veldkampi	Pteropodidae	5	0	0	0	0
Total		7404	59[0.8]	373[5.0]	651[8.8]	209[2.8]

Do demographic factors influence coronavirus infection risk?

Age was present in all but one of the six top model sets (table 2, figure 2a). It was a moderate to strong predictor of infection for four sets, and a weak predictor in one (table 2). Juveniles were more likely to be infected than adults, with odds ratios ranging from 1.58 to 5.68. Males *H.* cf. *ruber* appeared more likely to be infected with Alpha229E-CoV than females, however it was not important in the remaining top model sets (table 2, figure 2b). Reproductive status was included in two top-model sets, however for Alpha229E/*H. abae* this probably results from the higher prevalence among juveniles (table 2, figure 2c). For BetaBII, non-reproductive males and post-lactating females had higher infection probabilities, although the lower confidence intervals were close to one and should therefore be interpreted with caution.

Temporal patterns in CoV infections of West African bats

Month of capture was included in five of six top model sets (table 2). The highest statistically supported odds ratios occurred between April and July in each case, and in four of the five cases occurred between April and May (table 2, figure 3). This corresponds with the major wet season, as well as the primary birthing and nursing period for the three bat species. BetaC-CoV appeared

to have a second peak in September, correlated with the minor wet season and a second, smaller birthing pulse, however sample sizes were very low and confidence intervals included 1. Similarly, Alpha229E-CoV had peaks in the major and minor wet seasons for *H. abae* that were not statistically supported. There was an observable trend of decreased prevalence of BetaBII virus from mid 2011 to 2012. We see some evidence of yearly variation in BetaBII (increasing from 2010 to 2012) and BetaBI virus (decreasing 2011 to 2012).

Table 2. (opposite) Multimodel parameter averages for predictive factors of CoV infection in bats. Variable importances and model parameters averaged over the top model sets (ΔAICc < 2). Av. OR = odds ratio. 95% CI = 95% confidence interval based on profile likelihood. RI = relative variable importance. NRF = reproductively inactive female; NRM = reproductively inactive male; RAM = reproductively active male; P = pregnant female; L = lactating female; PL = post-lactating female. † Calculated in a separate analysis of which included pregnant females and excluded the variable BMI. § present versus absent (ref = absent) ‡ Sex and Rep variables were calculated in separate candidate model sets (see methods text). — = Not included in the top model set. NA = not applicable or excluded due to low sample sizes. * indicates that the 95% CI does not include 1.

Table 2. Multimodel parameter averages for predictive factors of CoV infection in bats.

Predictor variable	BetaC (*N.* cf. *gambiensis*)		BetaBII (*H.* cf. *ruber*)		Alpha229E (*H.* cf. *ruber*)		Alpha229E (*H. abae*)		BetaBI (*H.* cf. *ruber*)		BetaBI (*H. abae*)	
	Av. OR (95%CI)	RI	Av. OR (95%CI)	RI	Av. OR (95%CI)	RI	Av. OR (95%CI)	RI	Av. OR (95%CI)	RI	Av. OR (95%CI)	RI
Month (ref=Jul)		0.98		0.98		1.00		—		1.00		0.99
Jan	0.9 (0.21—3.93)		0.68 (0.42—1.12)		1.29 (0.82—2.02)				0.29 (0.17—0.50)*		0.14 (0.01—1.40)	
Feb	—		—		0.18 (0.07—0.52)*				0.29 (0.09—0.96)*		—	
Mar	2.63 (0.74—9.4)		1.21 (0.71—2.05)		1.21 (0.72—2.03)				0.35 (0.19—0.63)*		1.48 (0.40—5.53)	
Apr	7.44 (1.76—31.51)*		0.99 (0.54—1.81)		3.40 (2.05—5.65)*				0.15 (0.06—0.33)*		0.86 (0.13—5.81)	
May	6.59 (1.44—30.13)*		0.36 (0.20—0.64)*		1.03 (0.71—1.49)				0.50 (0.35—0.72)*		2.86 (1.14—9.70)*	
Jun	0 (0—Inf)		0.48 (0.22—1.03)		0.68 (0.38—1.21)				0.34 (0.18—0.63)*		1.25 (0.18—8.63)	
Aug	—		—		0.16 (0.02—1.19)				0.51 (0.05—5.26)		—	
Sep	1.7e16 (0—Inf)		0.54 (0.29—1.01)		0.23 (0.10—0.54)*				0.14 (0.08—0.25)*		0.13 (0.01—1.26)	
Nov	—		0.80 (0.38—1.67)		0.94 (0.53—1.66)				0.32 (0.19—0.55)*		0.47 (0.05—4.80)	
Age Juv (ref=Ad)	5.68 (1.24—25.91)*	0.83		—	2.32 (1.61—3.33)*	1.00	3.09 (1.22—7.84)*	0.75	1.58 (1.21—2.06)*	0.98	2.26 (0.88—5.79)	0.53
Sex male (ref=F)‡				—	1.31 (1.01—1.70)*	0.74	1.38 (0.64—2.98)	0.33	1.10 (0.88—1.36)	0.34	1.57 (0.78—3.17)	0.42
BMI	0.83 (0.48—1.42)	0.32	0.92 (0.77—1.11)	0.35	0.51 (0.35—0.74)*	0.99	0.68 (0.36—1.28)	0.52	0.93 (0.68—1.27)	0.30	0.73 (0.42—1.26)	0.39
Rep (ref=NRF)		—		0.92		—		0.75				—
Juv			1.24 (0.60—2.57)				3.22 (1.01—10.25)*					
RAM			1.76 (0.97—3.18)				1.32 (0.30—5.78)					
NRM			1.84 (1.09—3.10)*				1.25 (0.43—3.63)					
P†			1.05 (0.39—1.28)				0.66 (0.11—3.86)					
L			0.64 (0.25—1.65)				0 (0—Inf)					
PL			2.53 (1.09—5.89)*				0 (0—Inf)					
Mites§		—	1.22 (0.80—1.84)	0.35	1.10 (0.78—1.56)	0.30	1.34 (0.49—3.67)	0.30	1.34 (0.99—1.82)	0.67		—
Streblids§		—	1.26 (0.89—1.78)	0.44	1.43 (1.10—1.86)*	0.92	0.62 (0.21—1.82)	0.36	1.14 (0.91—1.42)	0.41	0.83 (0.35—1.97)	0.28
Nycteribiids§		NA		NA	1.52 (0.79—2.90)	0.44		—	1.60 (0.91—2.80)	0.55		NA
Alpha229E-CoV		NA	1.17 (0.66—2.05)	0.30		NA		NA	1.86 (1.34—2.57)*	1.00		—
BetaBI-CoV		NA	1.38 (0.9—2.13)	0.50	1.95 (1.43—2.65)*	1.00	1.64 (0.6—4.5)	0.38		NA		NA
BetaBII-CoV		NA		—	1.14 (0.65—2.00)	0.29		—	1.37 (0.88—2.13)	0.47		NA

See description opposite

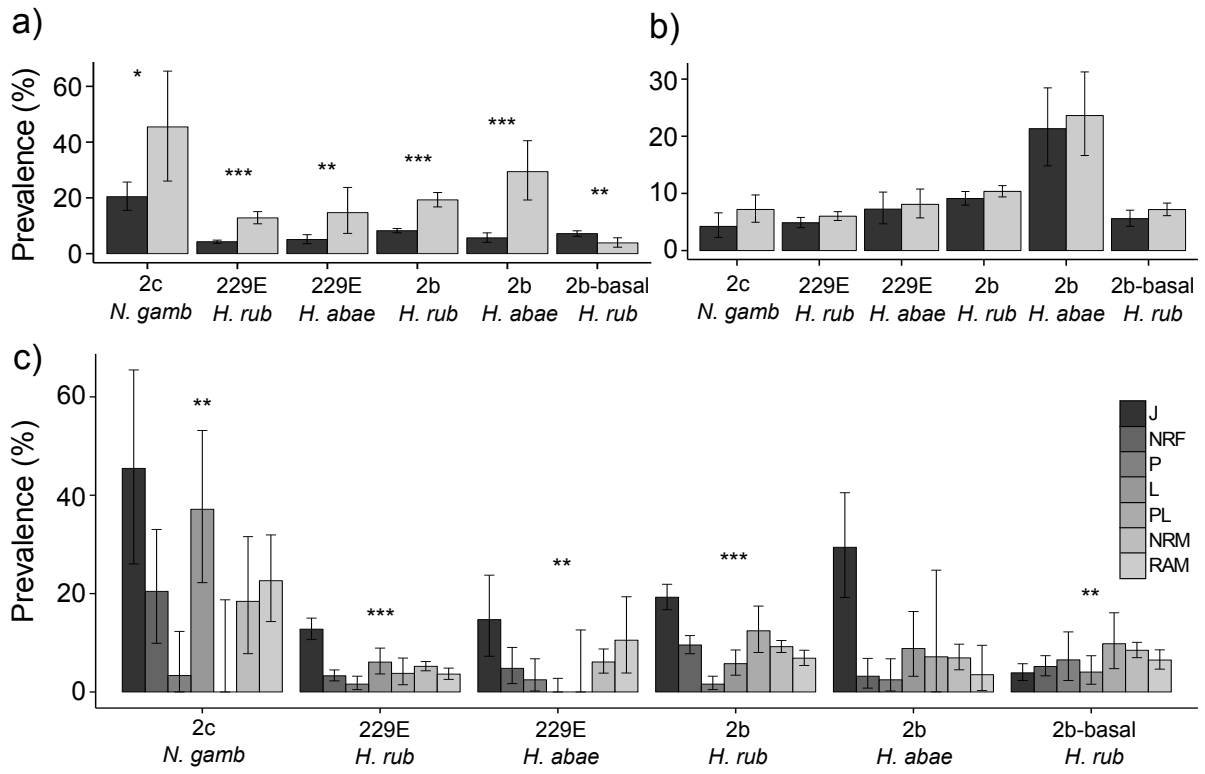


Figure 2. Prevalence as a function of age, sex and reproductive status across all sites, summarised throughout the study period. Pearson chi-square tests, * p<0.05, ** p<0.01, ***p<0.00001, blank = non-significant. a) Overall CoV prevalences by age. Error bars represent Bayesian binomial 95% CIs. Dark grey = adults, light grey = juveniles. b) Overall CoV prevalences by sex. Dark grey = female, light grey = male. c) Overall CoV prevalence by reproductive status. NRF = non-reproductive female, NRM = non-reproductive male, P = pregnant, L = lactating, PL = post-lactating, RAM = reproductively active male, J = juvenile.

Does body condition index predict infection risk?

Body condition index was included in all top model sets (table 2). Higher body condition index was associated with lower infection rates, however it was strongly predictive only for 229E-infections in *H.* cf. *ruber*, where an increase of one unit BMI was associated with 50% lower odds of infection.

Is co-infection with ectoparasites and other coronaviruses correlated with infection rate?

Infection with Alpha229E-CoV was associated with almost the doubled rate of BetaBI infection for *H.* cf. *ruber*, and vice-versa (table 2). Whilst other CoV pairs were positively correlated with one another in the top model sets, w^a values were low (0.47) and odds-ratios included 1.

There is some weak evidence that ectoparasites are associated with CoV infection risk. Streblid flies had a strong association with infection risk, but only for Alpha229E-CoV in *H.* cf. *ruber* (w^a = 0.92, OR = 1.43). Mites were positively associated with infection in four of the six top model sets, however the associations were considered unimportant (w^a < 0.67, CIs cross 1). Similarly, nycteribiids showed trends of positive association in two top model sets, however their influence was also unimportant (w^a < 0.55, CIs cross 1).

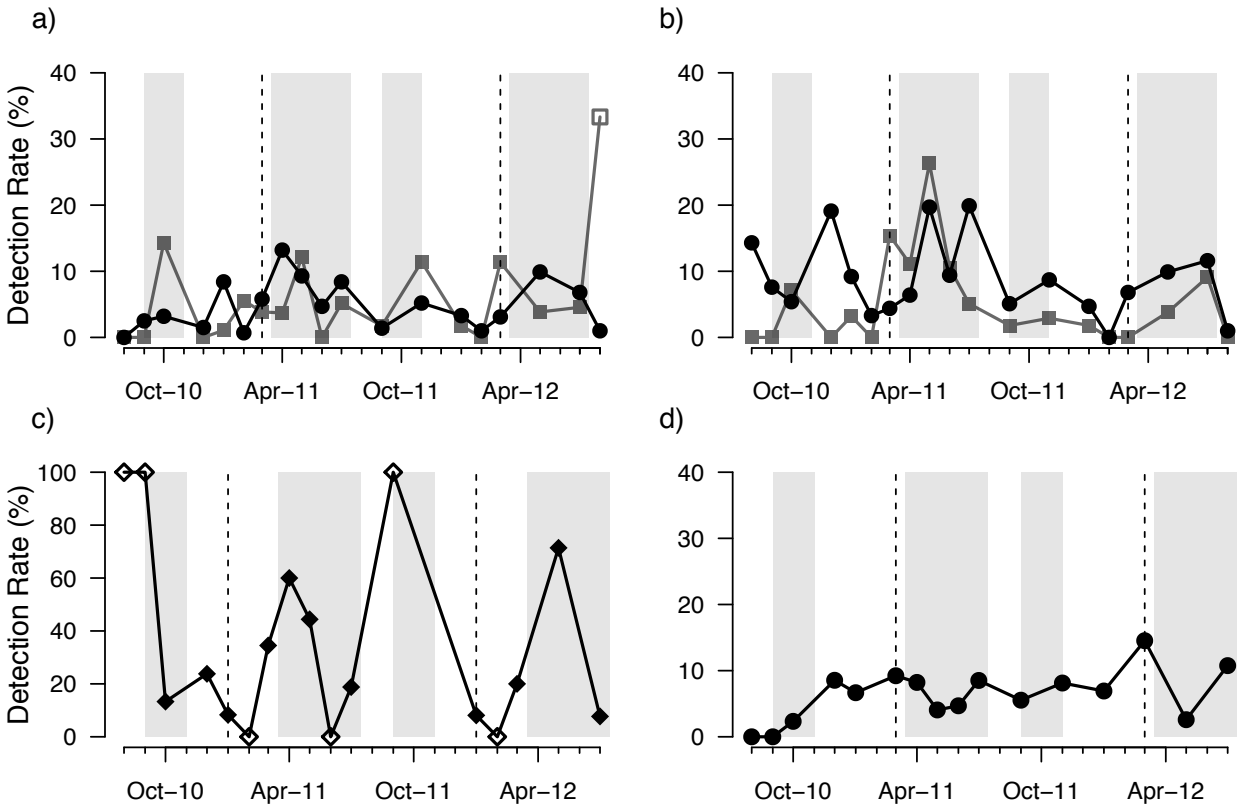


Figure 3. Proportion of CoV-infected animals caught at all sites summarised over the two years of the study. a) Alpha229E-CoV, b) BetaBI-CoV, c) BetaC-CoV, d) BetaBII-CoV. Circles = *H.* cf. *ruber*, squares = *H. abae*, triangles = *N.* cf. *gambiensis.* Open symbols indicate fewer than five samples per data point. Grey bars = rainy seasons. Dashed lines indicate birth of first pups – March for *H. abae* and *H.* cf. *ruber,* January for *Nycteris* cf. *gambiensis.*

DISCUSSION

This longitudinal study of CoVs in bat colonies sheds light on CoV dynamics and infection risk factors in wild hosts. We found infection by CoVs to be widespread in cave-dwelling bats, particularly among *Hipposideros* species. The CoVs in this study persisted from year to year, and vary in level of host restriction, as has been observed for this family (Drexler *et al.* 2014; Pfefferle *et al.* 2009; Tong *et al.* 2009).

Age class was a powerful predictor of infection for three of the four CoVs (BetaC, 229E and BetaBI). Juvenile bats were at substantially higher risk of infection than adults, suggesting an important role in virus transmission and maintenance. The role juveniles may play in virus prevalence has been observed in a preliminary study of BetaC-CoV in *N.* cf. *gambiensis* (Annan *et al.* 2013), studies of alpha CoVs in European and North American bats (Gloza-Rausch *et al.* 2008; Osborne *et al.* 2011), and MERS-CoV in dromedary camels (Meyer *et al.* 2014). A similar pattern of age-specific infection risk has been observed for rabies in vampire bats (Streicker *et al.* 2012) and big brown bats (George *et al.* 2011), and Hendra virus in little red flying-foxes (Plowright *et al.* 2008). Whether juvenile mortality is impacted by virus infection is unknown, however future studies may help to elucidate this by examining seroprevalence to compare past infection rates in adults with rates in juveniles.

A role of adult reproductive status in CoV infection risk was generally not well supported, being unimportant for three of the four CoVs. However, there was some evidence of increased risk of BetaBII infection for non-reproductive males and post-lactating females, suggesting that this virus may have different transmission dynamics warranting further investigation. Interestingly, a previous finding of higher risk of BetaC-CoV infection for lactating females than non-lactating females (Annan *et al.* 2013) was not statistically supported by our results comparing reproductive statuses for males and females, suggesting that this was likely an artifact of a seasonally higher prevalence. This highlights the importance of a multivariable approach.

Risk of CoV infection was higher at particular times of the year, with season an important

predictor even after age class was taken into account. Seasonal factors can have an important influence on disease dynamics (Altizer *et al.* 2006), and have been observed for other bat viruses, including other corona-, filo-, henipa-, astro- and lyssaviruses (Amman *et al.* 2012; Drexler *et al.* 2011; George *et al.* 2011; Osborne *et al.* 2011; Plowright *et al.* 2008; Serra-Cobo *et al.* 2013; Wacharapluesadee *et al.* 2010). In this study, prevalence tended to peak around the time of the major birthing and nursing pulse for the three bat species. This suggests that horizontal transfer among juveniles may be key to long-term maintenance of CoVs, whereby they exploit a seasonal influx of immunologically naïve individuals, promoting a seasonal peak in virus prevalence which then affects the whole colony. We observed highest significant peaks in prevalence in the period between April and July, although there were also prevalence spikes outside of these times. This directly follows the birth of the first pups of *H. abae* and *H.* cf. *ruber* in March and the strongest birth pulse for *N.* cf. *gambiensis* (unpublished data), which generally occurs from April to July. This is consistent with previous work by Drexler *et al.* (2011), which showed higher CoV amplification from a European *Myotis myotis* colony during the two months following the birth of juveniles. These seasonal pulses may confer a higher risk of spillover to humans, as demonstrated by Amman *et al.* (2012) for Marburg virus infection in *Rousettus aegyptiacus.*

A further explanation for seasonal differences may be related to fluctuating cave occupancy. Although roosts are occupied year-round, a yearly drop in population density was observable towards end of the major wet season in July, particularly pronounced for *H.* cf. *ruber*, which far outnumber the other species at all sites (authors' observations). Such seasonal fluctuations in cave occupancy may influence contact rates between individuals driving temporal fluctuations in virus transmission rates (Drexler *et al.* 2011).

We found correlations with likelihood of virus infection for both microparasites and ectoparasites. CoV co-infection was observed in *H.* cf. *ruber* and *H. abae*. Our models showed strong correlation between BetaBI and Alpha229E viruses in *H. ruber*, with weaker evidence for a correlation between BetaBII- and BetaBI CoVs in *H.* cf. *ruber*, and consistent (though

unsupported) trends for the two *H. abae* CoVs. It is possible that an increased sample size for *H. abae* may show stronger predictive power of co-infection with these two viruses, or perhaps immunological or ecological differences between these species influence this variation. Our assays detect only a fraction of the potential pathogen microbiome, and future broader microbiome studies facilitated by next-generation sequencing technology may provide further insights.

Ectoparasites were weak to strong predictors of infection. Where variable importance was higher than 0.4, associations were positive, however these were strongly supported only for *H.* cf. *ruber*, with streblids as a predictor of infection for Alpha229E-CoV and mites and nycteribiids for BetaBI infection. These findings suggest that susceptibility to CoV infection may be increased by other CoV- or ectoparasitic co-infections, or may be caused by inherent differences between hosts. A study of time-series data on microparasitic infections in field voles found that parasitic infections and interactions drive infection risk, even when host and environmental variabilities are accounted for (Telfer *et al.* 2010). Parasitic infections can have both positive and negative effects on other infections (Hawley and Altizer 2011; Telfer *et al.* 2010). This may be caused by altered efficiency of immune system mechanisms, such as inhibition, or augmentation of, cytokine responses (Bordes and Morand 2011; Telfer *et al.* 2010).

Although little is known about the effects of CoV or macroparasitic infections on bat health, co-infection from micro- or macroparasites may exacerbate morbidity and mortality (Bordes and Morand 2011) and alter the length of infections or transmission efficiency (Rodriguez *et al.* 1999; Telfer *et al.* 2010), potentially influencing disease dynamics (Gay *et al.* 2014). Polyparasitism may be a key characteristic of host species involved in infectiousness, and knowledge of its prevalence may therefore improve understanding of disease dynamics and the mechanisms underpinning emerging infectious diseases (Bordes and Morand 2011; Gay *et al.* 2014).

Finally**,** higher BMI was strongly associated with a 50% lower CoV infection for Alpha229E infections in *H.* cf. *ruber*, with a negative correlation observed in all other cases. This

may provide the first indication that a CoV may negatively impact bat health. Alternatively, it may indicate that nutritional stress may increase susceptibility to CoVs. Low BMI or nutritional stress may negatively affect the immune system of bats, perhaps due to substantial energetic costs of immune function, which may allow increased viral replication (Sheldon and Verhulst 1996). Nutritional stress was found to lead to higher Hendra virus seroprevalence in little red flying foxes (Plowright *et al.* 2008). If nutritional stress indeed causes increased susceptibility, then other factors known to alter food availability and lower body condition, such as habitat degradation, fragmentation, roost disturbance, urbanization, drought or climate change have the potential to increase virus prevalence and the likelihood of spillover events, and may therefore drive disease emergence (Plowright *et al.* 2008).

Many important questions remain about the ecology and maintenance of coronaviruses in bats, and there may be other relevant factors in CoV infection risk beyond the scope of this study. Roosting ecology and species associations (Gay *et al.* 2014), spatial dynamics, population or individual genetic factors (Guivier *et al.* 2011; Meyer-Lucht and Sommer 2009), social behaviour (Woodroffe *et al.* 2009) and habitat fragmentation and degradation may all play roles in wildlife disease dynamics. Investigation of potential morbidity or mortality caused by CoV infection on host bats will also enhance our understanding of transmission dynamics and long-term maintenance (Streicker *et al.* 2012). There is a pressing need for more insight into the reservoirs, virus-host dynamics, and drivers of maintenance and spillover events of wildlife diseases, as part of a One Health approach to the prediction and prevention of zoonotic disease outbreaks.

In this paper, we provide evidence that seasonal processes, demographic and ecological factors influence the risk of CoV infection in bats. These findings carry real-world implications for public health management and prevention of zoonotic disease emergences. We focused on bats in West Africa, a region with high direct and indirect human-bat contact (Anti *et al.* 2015). Bats are commonly consumed as bush meat in West Africa (Mickleburgh *et al.* 2009) and from some of our study sites (Anti *et al.* 2015), and bats including *N. gambiensis* and *Hipposideros* spp.

are known to roost in buildings in direct contact with humans or livestock (personal observations). In Ghana, visitation of bat caves for spiritual reasons is common, and collection of resources such as water from bat caves for household use, and occasionally for personal consumption, also occurs (authors' personal observations). Based on our findings, avoiding consumption of juvenile bats, and avoiding cave visitation, resource collection and bat consumption during at least April to May, may be beneficial in reducing the likelihood of human exposure to bat CoVs. This highlights some of the deep cultural and socioeconomic challenges in the prevention of emerging infectious diseases. Interventions such as community awareness and education are important, but may see limited effectiveness if fundamental issues such as poverty, food security and access to resources such as clean water are not also included in the global efforts to prevent emerging infectious diseases.

Ethics statement

Permission for capturing and sampling was obtained from the Wildlife Division of the Forestry Commission of the Ministry of Lands, Forestry and Mines, under research permit A04957 and ethics permit CHRPE49/09/CITES. All animals were handled according to the European Union Council Directive 86/609/EEC for the protection of animals and in accordance with Ghanaian law.

ACKNOWLEDGEMENTS

We thank Sebastian Brünink, Tobias Bleicker, and Monika Eschbach-Bludau for technical assistance. We are grateful to Isaac Mawusi Adanyeguh, Lucinda Kirkpatrick, Anna Vogeler, Mac Elikem Nutsuakor, David Ofori Agyei, Sarah Koschnicke, Julia Morrison, Emmanual Asare, Eunice Okyere, Kenneth Quansah, Emmanuel Essoun, Julian Schmid, Florian Gloza-Rausch, Maximilian Vollstaedt, Ebenezer Gyimah, Paul Marfo, Kennedy Darkwa, Justice Konadu, Paul, Kwame Takye, Winfred, Samuel, Francis and Thomas Kruppa for their help during the organization and conduct of field work. We thank Mirjam Knörnschild and Simone Sommer for

helpful discussion, and Karina Montero and Simon Ripperger for comments on early versions of the manuscript. We thank the communities, community leaders and administrators of the sampling sites, the helpful staff of the Ghana Wildlife Division, our local guides, and KCCR support staff.

REFERENCES

Altizer S., Dobson A., Hosseini P., Hudson P., Pascual M. & Rohani P. (2006) Seasonality and the dynamics of infectious diseases. *Ecology Letters* 9(4), 467-484.

Amman B. R., Carroll S. A., Reed Z. D., Sealy T. K., Balinandi S., Swanepoel R., Kemp A., Erickson B. R., Comer J. A. & Campbell S. (2012) Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. *PLOS Pathogens* 8(10), e1002877.

Annan A., Baldwin H., Corman V., Klose S., Owusu M., Nkrumah E., Badu E., Anti P., Agbenyega O., Meyer B., Oppong S., Sarkodie Y., Kalko E., Lina P., Godlevska E., Reusken C. B. E. M., Seebens A., Gloza-Rauch F., Vallo P., Tschapka M., Drosten C. & Drexler J. F. (2013) Human betacoronavirus 2c EMC/2012–related viruses in bats, Ghana and Europe. *Emerging Infectious Diseases* 19(3), 456-459.

Anti P., Owusu M., Agbenyega O., Annan A., Badu E. K., Nkrumah E. E., Oppong S. K., Adu-Sarkodie Y. & Drosten C. (2015) Human–bat interactions in rural West Africa. *Emerging Infectious Diseases* 21(8), 1418-1421.

Bartoń K. (2013) MuMIn: Multi-model inference. R package version 1.9.13. URL: http://CRAN.R-project.org/package=MuMIn.

Bates D., Mächler M., Bolker B. & Walker S. (2013) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5. URL http://CRAN.R-project.org/package=lme4.

Bordes F. & Morand S. (2011) The impact of multiple infections on wild animal hosts: a review. *Infection Ecology & Epidemiology* 1, 7346-7355.

Bradley C. A. & Altizer S. (2006) Urbanization and the ecology of wildlife diseases. *Trends in Ecology & Evolution* 22(2), 95-102.

Brunet-Rossinni A. K. & Wilkinson G. S. (2009) Methods for age estimation and the study of senescence in bats. In: *Ecological and Behavioral Methods for the Study of Bats* (eds T. H. Kunz and S. Parsons) pp. 315-325. Johns Hopkins University Press, Baltimore.

Burnham K. P. & Anderson D. R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York

Calisher C. H., Childs J. E., Field H. E., Holmes K. V. & Schountz T. (2006) Bats: important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews* 19(3), 531-545.

Corman V., Eickmann M., Landt O., Bleicker T., Brunink S., Eschbach-Bludau M., Matrosovich M., Becker S. & Drosten C. (2013a) Specific detection by real-time reverse-transcription PCR assays of a novel avian influenza A (H7N9) strain associated with human spillover infections in China. *Eurosurveillance* 18(16), 20461.

Corman V. M., Rasche A., Diallo T. D., Cottontail V. M., Stöcker A., Souza B. F. d. C. D., Corrêa J. I., Carneiro A. J. B., Franke C. R., Nagy M., Metz M., Knörnschild M., Kalko E. K. V., Ghanem S. J., Morales K. D. S., Salsamendi E., Spínola M., Herrler G., Voigt C. C., Tschapka M., Drosten C. & Drexler J. F. (2013b) Highly diversified coronaviruses in neotropical bats. *Journal of General Virology* 94, 1984-1994.

Daszak P., Cunningham A. & Hyatt A. (2000) Emerging infectious diseases of wildlife- threats to biodiversity and human health. *Science* 287(5452), 443-449.

Daszak P., Cunningham A. & Hyatt A. (2001) Anthropogenic environmental change and the emergence of infectious diseases in wildlife. *Acta Tropica* 78(2), 103-116.

Dobson A. P. (2005) What links bats to emerging infectious diseases? *Science* 310(5748), 628-629.

Drexler J. F., Corman V. M. & Drosten C. (2014) Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antiviral Research* 101(0), 45-56.

Drexler J. F., Corman V. M., Muller M. A., Maganga G. D., Vallo P., Binger T., Gloza-Rausch F., Rasche A., Yordanov S., Seebens A., Oppong S., Sarkodie Y. A., Pongombo C., Lukashev A. N., Schmidt-Chanasit J., Stocker A., Carneiro A. J. B., Erbar S., Maisner A., Fronhoffs F., Buettner R., Kalko E. K. V., Kruppa T., Franke C. R., Kallies R., Yandoko E. R. N., Herrler G., Reusken C., Hassanin A., Kruger D. H., Matthee S., Ulrich R. G., Leroy E. M. & Drosten C. (2012) Bats host major mammalian paramyxoviruses. *Nature Communications* 3, 796.

Drexler J. F., Corman V. M., Wegner T., Tateno A. F., Zerbinati R. M., Gloza-Rausch F., Seebens A., Müller M. A. & Drosten C. (2011) Amplification of emerging viruses in a bat colony. *Emerging Infectious Diseases* 17(3), 449-456.

Drexler J. F., Gloza-Rausch F., Glende J., Corman V. M., Muth D., Goettsche M., Seebens A., Niedrig M., Pfefferle S. & Yordanov S. (2010) Genomic characterization of SARS-related coronavirus in European bats and classification of coronaviruses based on partial *RNA-dependent RNA polymerase* gene sequences. *Journal of Virology* 84(21), 11336-11349.

Drosten C. (2013) Virus ecology: a gap between detection and prediction. *Emerging Microbes & Infections* 2(5), e31.

Fox J. & Monette G. (1992) Generalized collinearity diagnostics. *Journal of the American Statistical Association* 87(417), 178-183.

Fox J. & Weisberg S. (2011) *An R Companion to Applied Regression*. Sage, Thousand Oaks CA.

Gay N., Olival K. J., Bumrungsri S., Siriaroonrat B., Bourgarel M. & Morand S. (2014) Parasite and viral species richness of Southeast Asian bats: fragmentation of area distribution matters. *International Journal for Parasitology: Parasites and Wildlife* 3(2), 161-170.

Geldenhuys M., Weyer J., Nel L. H. & Markotter W. (2013) Coronaviruses in South African bats. *Vector-Borne and Zoonotic Diseases* 13(7), 516-519.

George D., Webb C., Farnsworth M., O'Shea T., Bowen R., Smith D., Stanley T., Ellison L. & Rupprecht C. (2011) Host and viral ecology determine bat rabies seasonality and

maintenance. *Proceedings of the National Academy of Sciences of the United States of America* 108(25), 10208-10213.

Gloza-Rausch F., Ipsen A., Seebens A., Göttsche M., Panning M., Drexler J., Petersen N., Annan A., Grywna K. & Müller M. (2008) Detection and prevalence patterns of group I coronaviruses in bats, northern Germany. *Emerging Infectious Diseases* 14(4), 626-631.

Guivier E., Galan M., Chaval Y., Xuéreb A., Ribas Salvador A., Poulle M. L., Voutilainen L., Henttonen H., Charbonnel N. & Cosson J. F. (2011) Landscape genetics highlights the role of bank vole metapopulation dynamics in the epidemiology of Puumala hantavirus. *Molecular Ecology* 20(17), 3569-3583.

Haagmans B. L., Al Dhahiry S. H. S., Reusken C. B. E. M., Raj V. S., Galiano M., Myers R., Godeke G.-J., Jonges M., Farag E., Diab A., Ghobashy H., Alhajri F., Al-Thani M., Al-Marri S. A., Al Romaihi H. E., Al Khal A., Bermingham A., Osterhaus A. D. M. E., AlHajri M. M. & Koopmans M. P. G. (2014) Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *The Lancet Infectious Diseases* 14(2), 140-145.

Hawley D. M. & Altizer S. M. (2011) Disease ecology meets ecological immunology: understanding the links between organismal immunity and infection dynamics in natural populations. *Functional Ecology* 25, 48-60.

Hegyi G. & Garamszegi L. Z. (2011) Using information theory as a substitute for stepwise regression in ecology and behavior. *Behavioral Ecology and Sociobiology* 65(1), 69-76.

Ithete N. L., Stoffberg S., Corman V. M., Cottontail V. M., Richards L. R., Schoeman M. C., Drosten C., Drexler J. F. & Preiser W. (2013) Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerging Infectious Diseases* 19(10), 1697-1699.

Jones K., Patel N., Levy M., Storeygard A., Balk D., Gittleman J. & Daszak P. (2008) Global trends in emerging infectious diseases. *Nature* 451(7181), 990-993.

Kunz T. H., Braun de Torrez E., Bauer D., Lobova T. & Fleming T. H. (2011) Ecosystem services provided by bats. *Annals of the New York Academy of Sciences* 1223(1), 1-38.

Lau S., Woo P., Li K., Huang Y., Tsoi H., Wong B., Wong S., Leung S., Chan K. & Yuen K. (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences of the United States of America* 102(39), 14040.

Leroy E. M., Epelboin A., Mondonge V., Pourrut X., Gonzalez J.-P., Muyembe-Tamfum J.-J. & Formenty P. (2009) Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector-Borne and Zoonotic Diseases* 9(6), 723-728.

Li W., Shi Z., Yu M., Ren W., Smith C., Epstein J., Wang H., Crameri G., Hu Z., Zhang H., Zhang J., McEachern J., Field H., Daszak P., Eaton B. T., Zhang S. & Wang L. F. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310(5748), 676-679.

Luis A. D., Hayman D. T. S., O'Shea T. J., Cryan P. M., Gilbert A. T., Pulliam J. R. C., Mills J. N., Timonin M. E., Willis C. K. R., Cunningham A. A., Fooks A. R., Rupprecht C. E., Wood J. L. N. & Webb C. T. (2013) A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proceedings of the Royal Society of London B Biological Sciences* 280(1756), 20122753.

Meyer B., Müller M. A., Corman V. M., Reusken C., Ritz D., Godeke G., Lattwein E., Kallies S., Siemens A. & van Beek J. (2014) Antibodies against MERS coronavirus in dromedary camels, United Arab Emirates, 2003 and 2013. *Emerging Infectious Diseases* 20(4), 552-559.

Meyer-Lucht Y. & Sommer S. (2009) Number of MHC alleles is related to parasite loads in natural populations of yellow necked mice, *Apodemus flavicollis*. *Evolutionary Ecology Research* 11(7), 1085-1097.

Mickleburgh S., Waylen K. & Racey P. (2009) Bats as bushmeat: a global review. *Oryx* 43(2), 217-234.

Osborne C., Cryan P. M., O'Shea T. J., Oko L. M., Ndaluka C., Calisher C. H., Berglund A. D., Klavetter M. L., Bowen R. A. & Holmes K. V. (2011) Alphacoronaviruses in New World bats: prevalence, persistence, phylogeny, and potential for interaction with humans. *PLOS One* 6(5), e19156.

Peig J. & Green A. J. (2009) New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. *Oikos* 118(12), 1883-1891.

Pfefferle S., Oppong S., Drexler J., Gloza-Rausch F., Ipsen A., Seebens A., Müller M., Annan A., Vallo P., Adu-Sarkodie Y., Kruppa T. & Drosten C. (2009) Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerging Infectious Diseases* 15(9), 1377-1384.

Plowright R., Field H., Smith C., Divljan A., Palmer C., Tabor G., Daszak P. & Foley J. (2008) Reproduction and nutritional stress are risk factors for Hendra virus infection in little red flying foxes (*Pteropus scapulatus*). *Proceedings of the Royal Society of London B Biological Sciences* 275(1636), 861-869.

Quan P.-L., Firth C., Street C., Henriquez J. A., Petrosov A., Tashmukhamedova A., Hutchison S. K., Egholm M., Osinubi M. O. & Niezgoda M. (2010) Identification of a severe acute respiratory syndrome coronavirus-like virus in a leaf-nosed bat in Nigeria. *MBio* 1(4), e00208-00210.

R Core Team. (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rodriguez M., Terrazas L. I., Marquez R. & Bojalil R. (1999) Susceptibility to *Trypanosoma cruzi* is modified by a previous non‐related infection. *Parasite Immunology (Oxford)* 21(4), 177-185.

Saéz A. M., Weiss S., Nowak K., Lapeyre V., Zimmermann F., Düx A., Kühl H. S., Kaba M., Regnaut S. & Merkel K. (2015) Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Molecular Medicine* 7(1), 17-23.

Saif L. (2004) Animal coronaviruses: what can they teach us about the severe acute respiratory syndrome? *Revue scientifique et technique (International Office of Epizootics)* 23(2), 643-660.

Serra-Cobo J., López-Roig M., Seguí M., Sánchez L. P., Nadal J., Borrás M., Lavenir R. & Bourhy H. (2013) Ecological factors associated with European bat lyssavirus seroprevalence in Spanish bats. *PLOS One* 8(5), e64467.

Sheldon B. C. & Verhulst S. (1996) Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. *Trends in Ecology & Evolution* 11(8), 317-321.

Sing T., Sander O., Beerenwinkel N. & Lengauer T. (2005) ROCR: visualizing classifier performance in R. *Bioinformatics* 21(20), 3940-3941.

Streicker D. G., Recuenco S., Valderrama W., Gomez Benavides J., Vargas I., Pacheco V., Condori Condori R. E., Montgomery J., Rupprecht C. E., Rohani P. & Altizer S. (2012) Ecological and anthropogenic drivers of rabies exposure in vampire bats: implications for transmission and control. *Proceedings of the Royal Society of London B Biological Sciences* 279(1742), 3384-3392.

Swanepoel R., Smit S., Rollin P., Formenty P., Leman P., Kemp A., Burt F., Grobbelaar A., Croft J. & Bausch D. (2007) Studies of reservoir hosts for Marburg virus. *Emerging Infectious Diseases* 13(12), 1847-1851.

Tao Y., Tang K., Shi M., Conrardy C., Li K. S. M., Lau S. K. P., Anderson L. J. & Tong S. (2012) Genomic characterization of seven distinct bat coronaviruses in Kenya. *Virus Research* 167, 67-73.

Telfer S., Lambin X., Birtles R., Beldomenico P., Burthe S., Paterson S. & Begon M. (2010) Species interactions in a parasite community drive infection risk in a wildlife population. *Science* 330(6001), 243-246.

Tong S., Conrardy C., Ruone S., Kuzmin I. V., Guo X., Tao Y., Niezgoda M., Haynes L., Agwanda B. & Breiman R. F. (2009) Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerging Infectious Diseases* 15(3), 482-485.

Vallo P., Guillén-Servent A., Benda P., Pires D. & Koubek P. (2008) Variation of mitochondrial DNA in the *Hipposideros caffer* complex (Chiroptera: Hipposideridae) and its taxonomic implications. *Acta Chiropterologica* 10(2), 193-206.

Wacharapluesadee S., Boongird K., Wanghongsa S., Ratanasetyuth N., Supavonwong P., Saengsen D., Gongal G. & Hemachudha T. (2010) A longitudinal study of the prevalence of Nipah virus in *Pteropus lylei* bats in Thailand: evidence for seasonal preference in disease transmission. *Vector-Borne and Zoonotic Diseases* 10(2), 183-190.

Weiss S. R. & Navas-Martin S. (2005) Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiology and Molecular Biology Reviews* 69(4), 635-664.

Woo P. C. Y., Lau S. K. P., Huang Y. & Yuen K.-Y. (2009) Coronavirus diversity, phylogeny and interspecies jumping. *Experimental Biology and Medicine* 234(10), 1117-1127.

Woodroffe R., Donnelly C. A., Jenkins H. E., Johnston W. T., Cox D. R., Bourne F. J., Cheeseman C. L., Delahay R. J., Clifton-Hadley R. S., Gettinby G., Gilks P., Hewinson R. G., McInerney J. P. & Morrison W. I. (2006) Culling and cattle controls influence tuberculosis risk for badgers. *Proceedings of the National Academy of Sciences of the United States of America* 103(40), 14713-14717.

Woodroffe R., Donnelly C. A., Wei G., Cox D., Bourne F. J., Burke T., Butlin R. K., Cheeseman C., Gettinby G. & Gilks P. (2009) Social group size affects *Mycobacterium*

bovis infection in European badgers (*Meles meles*). *Journal of Animal Ecology* 78(4), 818-827.

World Health Organization. (2004) Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. http://www.who.int/csr/sars/country/table2004_2004_2021/en/.

Zaki A. M., van Boheemen S., Bestebroer T. M., Osterhaus A. D. M. E. & Fouchier R. A. M. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine* 367, 1814-1820.

Zhang G., Cowled C., Shi Z., Huang Z., Bishop-Lilly K. A., Fang X., Wynne J. W., Xiong Z., Baker M. L., Zhao W., Tachedjian M., Zhu Y., Zhou P., Jiang X., Ng J., Yang L., Wu L., Xiao J., Feng Y., Chen Y., Sun X., Zhang Y., Marsh G. A., Crameri G., Broder C. C., Frey K. G., Wang L.-F. & Wang J. (2013) Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* 339(6116), 456-460

SUPPLEMENTARY MATERIAL

Table S1. List of sites at which bats were captured and sampled for CoVs.

Site	Location	Site type	Latitude	Longitude	Sampling regime
Akpafu Todzi	AKP	mine	7.2619722	0.4915278	semiannual
Bobiri	BOB	open	6.6871	-1.34415	opportunistic
Botanical Gardens	KUM	open	6.6851111	-1.5618889	opportunistic
Bosumtwi	BOS	open	6.5395278	-1.4115278	opportunistic
Buoyem 1	BUO	cave	7.7235833	-1.9879167	bimonthly
Buoyem 2	BUO	cave	7.7238056	-1.9926389	bimonthly
Elmina	ELM	building	5.0827778	-1.3483056	semiannual
Forikrom	FOR	cave	7.58975	-1.8750833	bimonthly
KCCR	KUM	open	6.6698226	-1.5771767	opportunistic
Kwamang 1	KWM	cave	7.0035685	-1.3003098	bimonthly
Kwamang 2	KWM	cave	6.9832778	-1.2731944	bimonthly
Kwamang 3	KWM	open	7.0065315	-1.3012354	opportunistic
Likpe Todome 1	LIK	cave	7.1639444	0.6079167	semiannual
Likpe Todome 2	LIK	cave	7.1638611	0.6081389	semiannual
Shai Hills	SHA	cave	5.9290000	0.0750000	opportunistic

Mine = abandoned mine, open = open forested or farmed land

Table S2. Oligonucleotides used for PCR detection of bat coronaviruses in Ghana

Primer ID	Sequence (5' – 3')	Polarity
CoV-Hip-BetaBII-rtF	CAGGACGCRCTATTCGCTTA	+
CoV-Hip-BetaBII-rtP	JOE-CGAAGCGTAATGTGTTGCCCACCATAA-BHQ1	+(Probe)
CoV-Hip-BetaBII-rtR	TGCGCTTATAGCGTATTTCAAATT	-
BetaC-rtF	GCACTGTTGCTGGTGTCTCTATTCT	+
BetaC-rtP	JOE-TGACAAATCGCCAATACCATCAAAAGATGC-BHQ1	+(Probe)
BetaCrtR	GCCTCTAGTGGCAGCCATACTT	-
CoV-Hip-BetaBI-rt-F	TGCCTAATATGTTGCGTATTTTCG	+
CoV-Hip-BetaBI-rt-P	FAM-TCATTAATYTTGGCTCGTAAGCACTCGACG-BHQ1	+(Probe)
CoV-Hip-BetaBI-rt-R	ATARTATCGCTCACTCARCGTACAA	-
CoV-Alpha229E-F13948m	TCYAGAGAGGTKGTTGTTACWAAYCT	+
CoV-Alpha229E -P13990m	FAM-TGGCMACTTAATAAGTTTGGIAARGCYGG-BHQ1	+ (Probe)
CoV-Alpha229E -R14138m	CGYTCYTTRCCAGAWATGGCRTA	-

ID - identification; R=G/A, Y=C/T, S=G/C, W=A/T, M=A/C, K=G/T. PCR reactions were conducted using the SSIII RT-PCR Kit (Life Technologies, Karlsruhe, Germany). 25 µl reactions contained 5 µl of RNA, 12.5 µl reaction buffer, 1 µL enzyme mix , 0.4 µl of a 50 mM magnesium sulfate solution), 1µg of PCR-grade bovine serum albumin, 400 nM of each of the primers, as well as 100 nM of each probe. The thermal cycling profile involved 20 min at 50 °C for reverse transcription, followed by 3 min at 95 °C and 45 cycles of 15 sec at 95 °C, 20 sec at 56 °C and 10 sec at 72 °C.

Table S3. Global models of infection probability

Response variable	Global models
BetaC-CoV, *N.* cf. *gambiensis*	Mo + BMI + Mt + Str + Age + Sex
	Mo + BMI + Mt + Str + Rep
Alpha229E-CoV, *H.* cf. *ruber*	Mo + BMI + Mt + Str + Nyc + BetaBI + BetaBII + Age + Sex
	Mo + BMI + Mt + Str + Nyc + BetaBI + BetaBII + Rep
Alpha229E-CoV, *H. abae*	Mo + BMI + Mt + Str + BetaBI + Age + Sex
	Mo + BMI + Mt + Str + BetaBI + Rep
BetaBI-CoV, *H.* cf. *ruber*	Mo + BMI + Mt + Str + Nyc + Alpha229E + BetaBII + Age + Sex
	Mo + BMI + Mt + Str + Nyc + Alpha229E + BetaBII + Rep
BetaBI-CoV, *H. abae*	Mo + BMI + Mt + Str + Alpha229E + Age + Sex
	Mo + BMI + Mt + Str + Alpha229E + Rep
BetaBII-CoV, *H.* cf. *ruber*	Mo + BMI + Mt + Str + Alpha229E + BetaBI + Age + Sex
	Mo + BMI + Mt + Str + Alpha229E + BetaBI + Rep

Mo = month, BMI = body mass index, Rep = reproductive status, Mt = mites, Str = streblid flies, Nyc = Nycteribiid flies

Table S4. Results of CoV screening by sampling site

Site	Bat species	No. tested	No. positive			
			BetaC	Alpha229E	BetaBI	BetaBII
Kwamang 1	*Hipposideros abae*	279	0	13	29	1
	Hipposideros cf. *gigas*	13	0	0	0	0
	Hipposideros jonesi	39	0	0	1	0
	Hipposideros cf. *ruber*	1520	0	85	187	98
	Nycteris cf. *gambiensis*	12	0	0	0	0
	Rhinolophus landeri	2	0	0	0	0
Kwamang 2	*Hipposideros abae*	139	0	10	16	1
	Hipposideros cf. *gigas*	16	0	0	1	1
	Hipposideros jonesi	6	0	0	2	1
	Hipposideros cf. *ruber*	901	0	76	187	56
	Nycteris cf. *gambiensis*	133	30	0	0	0
	Rhinolophus landeri	3	0	0	0	0
Kwamang 3	*Hipposideros jonesi*	2	0	0	0	0
	Hipposideros cf. *ruber*	14	0	0	5	1
	Micropteropus pusillus	1	0	0	0	0
	Nanonycteris veldkampi	2	0	0	0	0
Buoyem 1	*Coleura afra*	25	0	0	0	0
	Nycteris cf. *gambiensis*	7	2	0	0	0
	Hipposideros abae	24	0	3	4	0
	Hipposideros cf. *ruber*	802	0	56	33	0
	Rousettus aegyptiacus	2	0	0	0	0
Buoyem 2	*Coleura afra*	82	0	0	0	0
	Hipposideros abae	202	0	5	8	0
	Hipposideros cf. *gigas*	12	0	0	0	0
	Hipposideros cf. *ruber*	1055	0	48	91	0
	Lissonycteris angolensis	7	0	0	0	0
	Nycteris cf. *gambiensis*	8	4	0	0	0
Forikrom	*Coleura afra*	47	0	0	0	0
	Hipposideros abae	83	0	6	2	0
	Hipposideros cf. *gigas*	8	0	0	0	0
	Hipposideros cf. *ruber*	699	0	39	53	50
	Lissonycteris angolensis	12	0	0	0	0
	Micropteropus pusillus	1	0	0	0	0
	Nycteris cf. *gambiensis*	101	19	0	0	0
Akpafu-Todzi	*Rhinolophus landeri*	11	0	0	0	0
	Hipposideros abae	54	0	9	0	0
	Hipposideros cf. *ruber*	628	0	7	4	0
	Lissonycteris angolensis	1	0	0	0	0
	Nycteris cf. *gambiensis*	4	0	0	0	0
	Rhinolophus alcyone	1	0	0	0	0
	Rhinolophus landeri	1	0	0	0	0
Likpe-Todome 1	*Hipposideros* cf. *ruber*	169	0	9	14	0
Likpe-Todome 2	*Hipposideros* cf. *ruber*	86	0	1	1	0
Elmina	*Hipposideros tephrus*	128	0	6	1	0
Bobiri	*Hipposideros fuliginosus*	1	0	0	0	0
	Hipposideros jonesi	2	0	0	0	0
	Hipposideros cf. *ruber*	2	0	0	0	0
	Myonycteris torquata leptodon	1	0	0	0	0
	Rhinolophus alcyone	4	0	0	0	0
Botanic Gardens	*Nanonycteris veldkampi*	3	0	0	0	0
	Myonycteris torquata	4	0	0	0	0
	Epomops buettikoferi	4	0	0	0	0
KNUST	*Epomops buettikoferi*	3	0	0	0	0
Lake Bosumtwi	*Micropteropus pusillus*	1	0	0	0	0
Shai Hills	*Coleura afra*	20	0	0	0	0
	Taphozous perforatus	21	0	0	0	0

Table S5. Top model sets (delta AICc<2)

Rank	Response variable and model structure	K	logLik	AICc	Δ_i	w_i	AUC
BetaC-CoV (n=157)							
1	Mo + Age	12	-75.99	178.14	0.00	0.22	0.75
2	Mo + Age + BMI	13	-75.74	180.03	1.89	0.08	0.76
Alpha229E-CoV, *H.* cf. *ruber* (n=4616)							
1	Mo + Age + Sex + BMI + Str + ßBI	19	-1002.17	2042.51	0.00	0.18	0.72
2	Mo + Age + Sex + BMI + Str + ßBI + Nyc	20	-1001.45	2043.08	0.58	0.13	0.73
3	Mo + Age + Sex + BMI + Str + ßBI + Mt	20	-1002.05	2044.29	1.79	0.07	0.73
4	Mo + Age + Sex + BMI + Str + ßBI + ßBII	20	-1002.06	2044.31	1.81	0.07	0.73
Alpha229E-CoV, *H. abae* (n=524)							
1	Rep	10	-109.34	239.12	0.00	0.06	0.71
2	Age	6	-113.54	239.24	0.13	0.05	0.69
3	Age + BMI	7	-112.82	239.86	0.75	0.04	0.66
4	Rep + BMI	11	-108.85	240.22	1.10	0.03	0.70
5	Rep + ßBI	11	-108.86	240.24	1.12	0.03	0.72
6	Rep + Str	11	-108.95	240.41	1.29	0.03	0.71
7	Age + ßBI	7	-113.11	240.44	1.32	0.03	0.66
8	Age + Str	7	-113.13	240.48	1.37	0.03	0.69
9	Age + Sex	7	-113.22	240.67	1.55	0.03	0.70
10	Mt + Rep	11	-109.19	240.90	1.79	0.02	0.73
11	Age + Mt	7	-113.38	240.99	1.87	0.02	0.69
12	Age+ BMI + Str	8	-112.38	241.05	1.93	0.02	0.66
13	Age+ BMI + ßBI	8	-112.41	241.09	1.98	0.02	0.66
BetaBI-CoV, *H.* cf. *ruber* (n=4616)							
1	Mo + Age + Mt + Nyc + α229	18	-1348.87	2733.90	0.00	0.03	0.77
2	Mo + Age + Mt + Nyc + α229 + ßBII	19	-1347.95	2734.06	0.16	0.03	0.78
3	Mo + Age + Mt + α229	17	-1350.07	2734.28	0.39	0.03	0.78
4	Mo + Age + Mt + α229 + ßBII	18	-1349.18	2734.51	0.61	0.02	0.78
5	Mo + Age + Mt + Str + Nyc + α229	19	-1348.28	2734.72	0.82	0.02	0.77
6	Mo + Age + Mt + Str + Nyc + α229 + ßBII	20	-1347.37	2734.92	1.02	0.02	0.78
7	Mo + Age + Mt + Str + α229	18	-1349.47	2735.08	1.18	0.02	0.77
8	Mo + Age + Sex + Mt + Nyc + α229	19	-1348.52	2735.20	1.31	0.02	0.77
9	Mo + Age + Mt + Str + α229 + ßBII	19	-1348.59	2735.35	1.45	0.02	0.78
10	Mo + Age + Sex + Mt + Nyc + α229 + ßBII	20	-1347.61	2735.40	1.51	0.02	0.78
11	Mo + Age + Nyc + α229	17	-1350.64	2735.41	1.51	0.02	0.77
12	Mo + Age + Nyc + α229 + ßBII	18	-1349.68	2735.51	1.61	0.01	0.77
13	Mo + Age + Sex + Mt + α229	18	-1349.72	2735.60	1.70	0.01	0.78
14	Mo + Age + BMI +Mt + Nyc + α229	19	-1348.77	2735.70	1.81	0.01	0.78
15	Mo + Age + Sex + Mt + α229 + ßBII	19	-1348.85	2735.87	1.97	0.01	0.78
16	Mo + Age + BMI + Mt + Nyc + α229 + ßBII	20	-1347.85	2735.88	1.99	0.01	0.78
BetaBI-CoV, *H. abae* (n=499)							
1	Mo + Age	13	-132.03	290.80	0.00	0.08	0.84
2	Mo + Age + Sex	14	-131.32	291.50	0.70	0.06	0.84
3	Mo + BMI	13	-132.53	291.81	1.00	0.05	0.84
4	Mo	12	-133.70	292.05	1.24	0.05	0.84
5	Mo + Sex + BMI	14	-131.74	292.35	1.55	0.04	0.85
6	Mo + Sex	13	-132.81	292.37	1.57	0.04	0.84
7	Mo + Age + BMI	14	-131.88	292.63	1.82	0.03	0.80
8	Mo + Age + Str	14	-131.94	292.75	1.94	0.03	0.84

Table S5. (cont.)

Rank	Response variable and model structure	K	logLik	AICc	Δ_i	w_i	AUC
BetaBII-CoV, *H.* cf. *ruber* (n=2533)							
1	Mo + Rep + ßBI	18	-617.98	1272.23	0.00	0.05	0.65
2	Mo + Rep	17	-619.12	1272.48	0.25	0.04	0.65
3	Mo + Rep + Str + ßBI	19	-617.12	1272.54	0.31	0.04	0.65
4	Mo + Rep + Str	18	-618.20	1272.67	0.44	0.04	0.64
5	Mo + Rep + Mt + ßBI	19	-617.42	1273.15	0.92	0.03	0.65
6	Mo + Rep + Mt	18	-618.57	1273.41	1.18	0.03	0.65
7	Mo + Rep + BMI + ßBI	19	-617.58	1273.47	1.24	0.02	0.65
8	Mo + Rep + BMI	18	-618.74	1273.75	1.52	0.02	0.65
9	Mo + Rep + BMI + Str + ßBI	20	-616.71	1273.76	1.53	0.02	0.65
10	Mo + Rep + Mt + Str + ßBI	20	-616.75	1273.84	1.61	0.02	0.65
11	Mo + Rep + BMI + Str	19	-617.82	1273.94	1.71	0.02	0.65
12	Mo + Rep + α229 + ßBI	19	-617.83	1273.96	1.73	0.02	0.65
13	Mo + Rep + Mt + Str	19	-617.85	1273.99	1.76	0.02	0.65
14	Mo + Rep + α229	18	-618.90	1274.07	1.84	0.02	0.65
15	Mo + Sex	13	-624.07	1274.29	2.05	0.02	0.64

Table S6. Fitted model parameters for highest ranked models

Predictor variable	BetaC (*N. gambiensis*)		BetaBII (*H. ruber*)		Alpha229E (*H. ruber*)		Alpha229E (*H. abae*)		BetaBI (*H. ruber*)		BetaBI (*H. abae*)	
	OR	**SE**	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**
Intercept	0.14	0.57	0.06	0.29	0.03	0.20	0.04	0.46	0.10	0.62	0.04	0.92
Month (ref=Jul)												
Jan	0.93	0.74	0.73	0.26	1.28	0.23			0.29	0.28	0.16	1.17
Feb					0.19	0.53			0.29	0.61		
Mar	2.64	0.65	1.37	0.28	1.21	0.26			0.35	0.31	1.64	0.66
Apr	7.46	0.74	1.25	0.33	3.40	0.26			0.15	0.42	0.82	0.96
May	6.68	0.78	0.39	0.29	1.03	0.19			0.50	0.19	2.52	0.63
Jun	1.61e-13	3.31e6	0.48	0.39	0.68	0.29			0.34	0.32	1.17	0.98
Aug		3.72e7			0.16	1.02			0.51	1.19		
Sep	1.02e16		0.61	0.34	0.23	0.44			0.14	0.29	0.13	1.16
Nov			0.84	0.39	0.94	0.29			0.32	0.27	0.47	1.18
Age Juv (ref=Ad)	5.99	0.75			2.31	0.19			1.86	0.13	2.36	0.45
Sex male (ref=F)‡					1.31	0.13						
BMI					0.51	0.19						
Rep (ref=NRF)												
Juv			1.22	0.37			3.45	0.57				
RAM			1.73	0.30			1.31	0.75				
NRM			1.81	0.26			1.25	0.54				
L			0.63	0.48			3.71e-10	2.88e4				
PL			2.51	0.43			2.50	7.56e3				
Mites§									1.34	0.15		
Streblids§					1.43	0.13						
Nycteribiids§									1.59	0.29		
Alpha229E-CoV									1.86	0.17		
BetaBI-CoV			1.41	0.22	1.95	0.16						
BetaBII-CoV												

OR = odds ratio, SE = standard error

Table S7. Fitted model parameters for largest models within top model sets

Predictor variable	BetaC (*N. gambiensis*)		BetaBII (*H. ruber*)		Alpha229E (*H. ruber*)		Alpha229E (*H. abae*)		BetaBI (*H. ruber*)		BetaBI (*H. abae*)	
	OR	**SE**	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**
Intercept	0.15	0.57	0.06	0.31	0.03	0.20	0.13	0.79	0.09	0.62	0.03	0.97
Month (ref=Jul)												
Jan	0.81	0.77	0.73	0.26	1.29	0.23			0.29	0.28	0.15	1.18
Feb					0.18	0.53			0.29	0.60		
Mar	2.59	0.65	1.42	0.28	1.21	0.26			0.35	0.31	1.71	0.66
Apr	7.40	0.73	1.31	0.33	3.44	0.26			0.15	0.41	1.01	0.98
May	6.68	0.77	0.40	0.30	1.03	0.19			0.50	0.19	2.67	0.62
Jun	1.21e-13	3.67e6	0.53	0.39	0.68	0.29			0.34	0.32	1.23	0.99
Aug					0.15	1.02			0.51	1.19		
Sep	6.77e16	6.71e7	0.59	0.34	0.23	0.44			0.14	0.29	0.13	1.16
Nov			0.81	0.39	0.94	0.29			0.32	0.27	0.51	1.18
Age Juv (ref=Ad)	4.93	0.81			2.32	0.19	2.57	0.48	1.59	0.13	2.27	0.46
Sex male (ref=F)‡					1.31	0.13					1.53	0.36
BMI	0.83	0.28	0.91	0.10	0.51	0.19	0.68	0.32				
Rep (ref=NRF)												
Juv			1.19	0.37								
RAM			1.83	0.31								
NRM			1.89	0.27								
L			0.67	0.49								
PL			2.61	0.43								
Mites§									1.33	0.15		
Streblids§			1.27	0.18	1.43	0.13	0.61	0.55	1.13	0.12		
Nycteribiids§					1.52	0.33			1.60	0.29		
Alpha229E-CoV									1.84	0.17		
BetaBI-CoV			1.41	0.22	1.95	0.16						
BetaBII-CoV									1.37	0.23		

OR = odds ratio, SE = standard error

CHAPTER 4
Isolation and characterization of 11 novel microsatellite loci in a West African leaf-nosed bat, *Hipposideros* aff. *ruber*

Image: Florian Gloza-Rausch

H. J. Baldwin, P. Vallo, M. G. Gardner, C. Drosten, M. Tschapka and A. J. Stow. (2014)

BMC Research Notes, 7:60

ABSTRACT

Background: Noack's leaf-nosed bat, *Hipposideros ruber*, is a cryptic species within the *Hipposideros caffer* species complex. Despite a widespread distribution in Africa and being host to potentially zoonotic viruses, the genetic structure and ecology of *H. ruber* is poorly known. Here we describe the development of 11 novel polymorphic microsatellite loci to facilitate the investigation of genetic structure.

Findings: We selected 20 microsatellite sequences identified from high throughput sequence reads and PCR amplified these for 38 individuals, yielding 11 consistently amplifying and scorable loci. The number of alleles per locus ranged from two to 12, and observed heterozygosities from 0.00 to 0.865. No evidence of linkage disequilibrium was observed, and nine of the markers showed no departure from Hardy-Weinberg equilibrium. We demonstrate successful amplification in two closely related species and two divergent lineages of the *H. caffer* species complex.

Conclusions: These new markers will provide a valuable tool to investigate genetic structure in the poorly understood *Hipposideros caffer* species complex.

Keywords: *Hipposideros ruber*, *Hipposideros caffer*, microsatellites, population genetics, Hipposideridae, bat

FINDINGS

Noack's leaf-nosed bat *Hipposideros ruber* (Noack, 1893) is one of two recognised cryptic species within the *Hipposideros caffer* (Sundevall, 1846) species complex. These bats are widespread throughout sub-Saharan Africa and among the most abundant mammals on the continent (Brosset 1984; Wright 2009). Mitochondrial evidence has shown the existence of several deeply divergent lineages within the *H. caffer* complex, which most likely constitute more than the two species (Vallo *et al.* 2008). Recently, they have been discovered to host viruses with zoonotic potential (Pfefferle *et al.* 2009), emphasizing the need for knowledge about their ecology in order

to gain insight into zoonotic processes and risk factors for public health. Microsatellites provide a powerful tool to investigate the poorly known ecology and life history of these bats, including genetic structure, social arrangements and mating systems. Assessment of nuclear gene flow through microsatellite analysis may thus help to shed light also on the taxonomy of this species complex. Microsatellites have been developed for several species in the *Hipposideros* genus (Echenique-Diaz *et al.* 2002; Guo *et al.* 2008; Liu *et al.* 2008). These markers represent, to our knowledge, the first set of microsatellite loci developed for African hipposiderid bats.

We isolated and characterized 11 microsatellite loci from a single, exclusively West African mitochondrial lineage of *Hipposideros ruber*, determined by sequencing of the cytochrome *b* gene (lineage D; Vallo *et al.* 2008). This lineage is henceforth called *H.* aff. *ruber* due to its distant evolutionary relationship to *H. ruber* s. str. from East Africa, and may represent a distinct species (Vallo *et al.* 2008). *Hipposideros* aff. *ruber* has been previously identified in central Ghana (Vallo *et al.* 2011) and seems to be the most abundant of the three main lineages of the *H. caffer* complex in this region (unpublished data). The markers described herein represent, to our knowledge, the first suite of microsatellites for an African hipposiderid bat.

DNA was extracted from wing tissue from eight individuals sampled from the Brong Ahafo and Volta regions in central Ghana. DNA was extracted using an innuPREP DNA mini kit (Analytik Jena, Jena, Germany). Five micrograms of pooled DNA from eight individuals was sent to AGRF (www.agrf.com.au), where high throughput sequencing was performed on a Roche GS FLX 454 sequencing machine as described elsewhere (Boomer and Stow 2010; Gardner *et al.* 2011; Margulies *et al.* 2005). QDD 1.3 (Meglécz *et al.* 2010) was used to screen for di- to hexanucleotide repeat motifs with a minimum of eight repeats. From the 1689 microsatellites identified, a total of 32 primer pairs flanking tetranucleotide repeats with 11-15 repeat motifs were designed using PRIMER3 (Rozen and Skaletsky 2000). Twenty primer pairs for which the annealing temperatures were most similar for each primer were selected for initial amplification trials. Amplification products from these primer pairs were visualised by electrophoresis on an

agarose gel, from which 13 pairs with strong, stutter-free amplification bands were selected for optimisation. Forward primers for these 13 loci were directly labelled with a fluorochrome at the 5' end. Twelve of these loci were successfully amplified by polymerase chain reaction (PCR), with one discarded due to the excessive amplification of non-specific product. PCR conditions for these 12 loci were optimized and genotyping was performed on 38 individuals (16 females, 22 males) sampled in Brong Ahafo and Volta Regions.

PCRs were performed using 10-50 ng of template DNA and reagent concentrations as follows: 200 µM each dNTP, one unit reaction buffer, between 2.0 and 2.5 mM $MgCl_2$, equal concentrations of forward and reverse primer (0.25–1.0 µM) and one unit *Taq* polymerase (see table 1). PCR amplification consisted of an initial denaturation at 94°C for 3 min followed by six touchdown cycles of 94°C denaturation for 30 s, annealing for 30 s with temperatures decreased by 2°C per cycle (55–47°C, 60–50°C, or 65–55°C; table 1), and polymerase extension step at 72°C for 45 s. Additional 35 cycles were conducted, of denaturation (94°C, 30 s), primer annealing (final touchdown temperature, 45 s), and polymerase extension (72°C, 45 s), followed by a final extension (72°C, 10 min). PCR products were electrophoresed using an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Allele sizes were determined via manual inspection using the software PEAK SCANNER 1.0 (Applied Biosystems), followed by automated binning performed using TANDEM 1.09 (Matschiner and Salzburger 2009). We reanalyzed 20% of individuals to evaluate data integrity. One locus (Hr3) was discarded due to high rounding error in the TANDEM analysis, indicating poor marker quality. MICRO-CHECKER 2.2.3 was used to assess the probability of scoring errors, allelic dropout and the presence of null alleles (van Oosterhout *et al.* 2004). No scoring errors or allelic dropout were detected, although there were potentially null alleles at loci Hr7 and Hr12. Locus Hr13 may also suffer from null alleles, though low allelic variability (one common and one rare allele) did not allow this to be confirmed (table 1).

Table 1. Characteristics and thermocycling conditions for 11 polymorphic microsatellites in the African leaf-nosed bat *Hipposideros* aff. *ruber*

Locus	Accession #	Repeat Motif	Primer sequences (5' – 3')	$MgCl_2/P_{FR}$	*Ta*(°C)	Size (bp)	N	N_A	H_O	H_E	HWE	P_{NULL}
Hr1	KM370156	$(GATA)_{13}$	F:TGGCAAGGTTAACACGAACC R:TCTCCCTCCCGCTCTTATCT	2.0mM/0.5µM	60-50	238-258	38	6	0.74	0.79	ns	0.028
Hr2	KM370157	$(TCTT)_{15}$	F:GAAGCACTGCTGGAAAGGTT R:GTTGAACTGGGTGGCCTTTA	2.0mM/0.25µM	60-50	311-339	34	8	0.77	0.76	ns	-0.009
Hr5	KM370160	$(GAAG)_{14}$	F:TGGGTGTTTCAGTTTCATGC R:TGGTCTATTTGTTTCCTTCCGTA	2.0mM/0.5µM	65-55	186-234	34	9	0.82	0.82	ns	-0.006
Hr6	KM370161	$(TCTT)_{13}$	F:GGGTTTCTTCAAATGTGTTTTC R:GCCTCCAAGACAAACAGAGG	2.0mM/0.5µM	55-47	204-240	37	8	0.70	0.73	ns	0.012
Hr7	KM370162	$(ATTT)_{11}$	F:AGCCAATGACAAGACTGCCTA R:CCAGTGAAGCAACGTCCTCT	2.0mM/0.5µM	65-55	144-172	33	8	0.42	0.68	***	0.173
Hr8	KM370163	$(ATCT)_{12}$	F:CTCAGCCCAAAGTCAAGGAG R:TGGCTATACGAATACAAAGATTAGACA	2.0mM/0.5µM	60-50	221-241	36	6	0.72	0.68	ns	-0.042
Hr9	KM370164	$(TCTA)_{12}$	F:TGCTATCTTCCATGAGGTCAGA R:TCTCTGTTGCTGAAGGAAAACTT	2.0mM/0.5µM	60-50	218-234	38	5	0.63	0.73	ns	0.061
Hr10	KM370165	$(TTAT)_{11}$	F:TCCACTGGAGTAAGAGATGTGTG R:GCACTGCAACAGTGAAAAGC	2.0mM/1.0µM	65-55	258-282	38	7	0.79	0.74	ns	-0.040
Hr11	KM370166	$(TTTC)_{14}$	F:CTCTTGCAATGAAGGCAATG R:CTGCCATGAGCTACCATGAG	2.0mM/0.5µM	65-55	106-154	37	12	0.87	0.86	ns	-0.018
Hr12	KM370167	$(GATA)_{12}$	F:TTGGTTTTCAGATCTTCTGGTG R:GAGTCTTCTGCCTGCTGGAC	2.5mM/0.5µM	60-50	277-293	38	4	0.42	0.60	**	0.140
Hr13	KM370168	$(TTTC)_{13}$	F:CCGAAGCCAATCTGGTTTTA R:GGGTCCTGCAGAAACACACT	2.0mM/1.0µM	65-55	321-329	34	2	0.00	0.06	ns	0.157

P_{FR} forward and reverse primer concentration, *Ta* annealing temperatures of touchdown cycles (see methods), N number of individuals, N_A number of alleles, H_O observed heterozygosity, H_E expected heterozygosity, HWE probability of deviation from Hardy-Weinberg equilibrium, P_{NULL} null allele frequency estimate (van Oosterhout), ns not significant., ** $p<0.01$, *** $p<0.001$

The program CERVUS was used to calculate number of alleles, observed (H_E) and expected (H_O) heterozygosities, and probabilities of identity (Kalinowski *et al.* 2007). All 11 loci were determined to be polymorphic, with a range of 2-12 alleles per locus (table 1). Tests for pairwise linkage disequilibrium and deviations from Hardy-Weinberg equilibrium with Bonferroni corrections were calculated using FSTAT 2.9.3 (Goudet 1995). Two loci (Hr7, Hr12) deviated significantly from the Hardy-Weinberg equilibrium with a homozygote excess (table 1). No linkage disequilibrium was detected between any loci. The probability of identity for the 11 loci was low at $1.6E^{-10}$ overall, and $8.7E^{-10}$ and $3.1E^{-9}$ for the Brong Ahafo and Volta localities, respectively. Probability of sibling identity was $1.4E^{-4}$, $2.2E^{-4}$ and $2.8E^{-4}$ for overall, Brong Ahafo and Volta, respectively.

In order to explore utility in closely related taxa, we tested whether these loci could be amplified across four related taxa in the genus *Hipposideros* using the PCR conditions specified above (table 2). All but one locus successfully amplified PCR product across the tested taxa.

These microsatellite loci provide useful resources for the study of population genetic structure of bats in the *Hipposideros caffer* complex, and likely also related species in this genus. These findings will help to address questions regarding connectivity, social behaviour, and zoonotic disease ecology in African leaf-nosed bats.

Table 2. Cross-amplification success in other *Hipposideros* species or lineages

Taxon	Hr1	Hr2	Hr5	Hr6	Hr7	Hr8	Hr9	Hr10	Hr11	Hr12	Hr13
H. abae	+	+	+	+	+	+	+	+	+	+	+
H. caffer tephrus	+	+	+	+	+	+	+	+	–	+	+
H. ruber (lin. B)[§]	+	+	+	+	+	+	+	+	+	+	+
H. ruber (lin. C)[§]	+	+	+	+	+	+	+	+	+	+	+

+ successful amplification with 1-2 bands visualised of expected size, – no PCR product observed

[§] sensu Vallo *et al.* 2008. lin. = lineage

Ethics statement

All animals were handled in accordance with Ghanaian legislation. Bat capture and sampling were authorized by permit from the Wildlife Division of the Ministry of Lands, Forestry and Mines in Ghana, and approved by the Macquarie University Ethics Committee. Exports were conducted under a state agreement between the Republic of Ghana and the Federal Republic of Germany, and to Australia with permission from the Department of Agriculture, Fisheries and Forestry.

Availability of the supporting data

The microsatellite sequences are available through the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov); GenBank accession numbers KM370156 – KM370168.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AS was responsible for the design of this study, supervision of the work and contributed to the interpretation of results. HB performed field sampling, data analysis and marker validation, and HB and PV drafted the manuscript. MG contributed to analysis of sequences. CD and MT coordinated field sampling and were responsible for the implementation of the study. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation and funds from Macquarie, Ulm and Bonn Universities. We would like to thank all those involved in fieldwork, in particular E. E.

Nkrumah, P. Anti, and E. K. Badu. We thank the communities of Buoyem and Likpe Todome for their hospitality and for providing access to the sites.

REFERENCES

Boomer J. & Stow A. (2010) Rapid isolation of the first set of polymorphic microsatellite loci from the Australian gummy shark, *Mustelus antarcticus* and their utility across divergent shark taxa. *Conservation Genetics Resources* 2(1), 393-395.

Brosset A. (1984) Chiroptères d'altitude du Mont Nimba (Guinée). Description d'une espèce nouvelle, *Hipposideros lamottei*. *Mammalia* 48(4), 545-556.

Echenique-Diaz L. M., Yokoyama J., Kawata M., Abe S. & Ishibashi Y. (2002) Isolation and characterization of microsatellite loci in the Bang's leaf‐nosed bat *Hipposideros turpis*. *Molecular Ecology Notes* 2(4), 396-397.

Gardner M. G., Fitch A. J., Bertozzi T. & Lowe A. J. (2011) Rise of the machines – recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* 11(6), 1093-1101.

Goudet J. (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* 86(6), 485-486.

Guo T., Hua P., Lin L. & Zhang S. (2008) Characterization of novel microsatellite loci in the great leaf-nosed bat, *Hipposideros armiger*, and cross-amplification in other related species. *Conservation Genetics* 9(4), 1063-1065.

Kalinowski S. T., Taper M. L. & Marshall T. C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16(5), 1099-1106.

Liu W., Hua P., Zhang J., Rossiter S. J. & Zhang S. (2008) Isolation and characterization of microsatellite loci in Pratt's Leaf-nosed Bat (*Hipposideros pratti*) and cross-species amplification in closely related taxa. *Conservation Genetics* 9(5), 1341-1343.

Margulies M., Egholm M., Altman W., Attiya S., Bader J., Bemben L., Berka J., Braverman M., Chen Y. & Chen Z. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057), 376-380.

Matschiner M. & Salzburger W. (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* 25(15), 1982-1983.

Meglécz E., Costedoat C., Dubut V., Gilles A., Malausa T., Pech N. & Martin J.-F. (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26(3), 403-404.

Pfefferle S., Oppong S., Drexler J., Gloza-Rausch F., Ipsen A., Seebens A., Müller M., Annan A., Vallo P., Adu-Sarkodie Y., Kruppa T. & Drosten C. (2009) Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerging Infectious Diseases* 15(9), 1377-1384.

Rozen S. & Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* 132(3), 365-386.

Vallo P., Benda P., Martínková N., Kaňuch P., Kalko E. K., Červený J. & Koubek P. (2011) Morphologically uniform bats *Hipposideros* aff. *ruber* (Hipposideridae) exhibit high mitochondrial genetic diversity in southeastern Senegal. *Acta Chiropterologica* 13(1), 79-88.

Vallo P., Guillén-Servent A., Benda P., Pires D. & Koubek P. (2008) Variation of mitochondrial DNA in the *Hipposideros caffer* complex (Chiroptera: Hipposideridae) and its taxonomic implications. *Acta Chiropterologica* 10(2), 193-206.

van Oosterhout C., Hutchinson W., Wills D. & Shipley P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4(3), 535-538.

Wright G. (2009) *Hipposideros caffer* (Chiroptera: Hipposideridae). *Mammalian Species* 845, 1-9.

CHAPTER 5
Genetic and acoustic diversification: cryptic speciation of Pliocene origin in Afrotropical bats (Hipposideridae: *Hipposideros*) in West Africa

H. J. Baldwin, P. Vallo, A. T. Ruiz, P. Anti, E. E. Ewald, E. K. Badu, E. K. V. Kalko, M. Tschapka and A. J. Stow.

Formatted for *Molecular Biology and Evolution*

ABSTRACT

Levels of biodiversity are globally underestimated, especially in tropical ecosystems. This is particularly so for bats compared to other vertebrate taxa, due to morphological conservatism. Here we investigate *Hipposideros caffer* and *H. ruber*, two morphologically similar, insectivorous bat species in the Afrotropics. From samples collected in Ghana, we evaluate cryptic speciation using nuclear genetic, acoustic and morphometric data. Further, we used Bayesian divergence dating to explore the 'species pump' and 'museum' hypotheses proposed to explain diversity in this region. We determine that four mitochondrial DNA clades represent distinct species. Microsatellite data showed no evidence of interbreeding between groups, and there are significant differences in acoustic and morphometric measurements. Divergence dating estimated that these four species and two close relatives arose in the Pliocene, with an origin of the broader group in the late Miocene, and intraspecific lineages corresponding to the Pleistocene. These findings are consistent with the 'species pump' hypothesis, which emphasizes an important role of climatic fluctuations in the species richness of forest refugia. Acoustic differences between species were consistent with social selection, drift or reproductive character displacement, but not resource partitioning, as mechanisms for the emergence and maintenance of acoustic divergence in these bats.

Keywords: *Hipposideros caffer, Hipposideros ruber*, bats, refugia

INTRODUCTION

The Afrotropic ecozone contains five of the world's 25 biodiversity hotspots (Myers *et al.* 2000). Despite this, poor governance, economic underdevelopment and a history of civil conflict have left its fauna understudied. Throughout the world, tropical forests are disappearing at an alarming rate, and several of Africa's biodiversity hotspots are also hotspots for deforestation (Malhi *et al.* 2013). Given this high diversity and the current catastrophic rate of rate of anthropogenic

extinction (Ceballos *et al.* 2015), there is a serious risk of losing species before they are even known to science. In order to make informed decisions about wildlife management, taxonomic resolution and knowledge of the processes that give rise to and maintain diversity are required.

Species richness of bats is exceptionally poorly known throughout much of Africa, and is likely to be far greater than currently described (Bates *et al.* 2013; Fahr and Kalko 2011; Monadjem *et al.* 2013b; Monadjem *et al.* 2010; Taylor *et al.* 2012). This poor state of knowledge may be compounded by a higher incidence of cryptic diversity in bats than other taxa (Jones and Barlow 2004). Species are considered cryptic if they share high morphological similarity and have been traditionally considered as a single species, but are actually distinct based on other lines of evidence (e.g. genetic, ecological, behavioural, subtle morphological; Bickford *et al.* 2007). In bats, morphological constraints due to flight and echolocation, and the relative importance of auditory versus visual mating signals, mean that divergence often goes undetected (Jones and Barlow 2004; Kingston *et al.* 2001). In recent years, molecular, acoustic, behavioural and detailed morphological analyses have revealed many cryptic bat species complexes (Jones 1997; Jones and Barlow 2004; Kingston *et al.* 2001; Monadjem *et al.* 2013a; Russo and Jones 2000; Thabah *et al.* 2006).

Accurate identification of species is important for rigorous assessments of biodiversity, species richness and a species' conservation status (Furman *et al.* 2010), allowing identification for prioritization and management (Moritz and Faith 1998). Where one recognised species of low conservation concern actually comprises two or more cryptic species, vulnerability of component taxa may be masked. Further, wildlife management for other purposes, such as the control or prevention of enzootic or zoonotic diseases, requires a detailed knowledge of host ecology in order to be effective. Taxonomic resolution is important for generating information about prevalence, susceptibility, risk factors, and spread of wildlife diseases, in order to design effective management strategies.

Knowledge of biodiversity patterns is also crucial for building a broader understanding of

the processes that generate and maintain diversity. While conservation management often focuses on patterns of diversity in order to maximise its representation, increasingly, the importance of preserving the processes that generate and maintain diversity is being recognised (Moritz 2002; Moritz and Faith 1998; Rouget *et al.* 2003). Preservation of evolutionary patterns and processes is critical for long-term population persistence, particularly in the face of the continuing threats posed to biodiversity by habitat destruction and degradation and climate change (Mace and Purvis 2008). It is therefore beneficial to improve our understanding of these processes, in order to make informed decisions with regards to managing their preservation.

High species diversity in the Afrotropics is often attributed to changes in the distribution and composition of forest induced by major shifts in the palaeoclimate (Hamilton and Taylor 1992; Maley 1989). The relatively warm, wet climate of the mid-Miocene gave way to a trend of slow global cooling and drying, punctuated by oscillations of wet and dry periods, and culminating in the Pleistocene glaciations (deMenocal 1995; 2004). This resulted in forest expansion during warmer, wetter periods, and contraction and replacement with savannah during the dry periods, with forest species contained within refugia (deMenocal 2004). Two broad hypotheses have been proposed to explain how refugia have shaped the current biodiversity in tropical Africa. The species pump hypothesis proposes that these climate fluctuations catalysed diversification, by creating new selective pressures, isolating populations in refugia and thereby promoting allopatric speciation (Murienne *et al.* 2013). This elevated speciation model is typically invoked through the paradigm of Pleistocene glaciations, however, the increased use of molecular markers has provided evidence for a substantial role of refugial processes extending back to the late Miocene and Pliocene (deMenocal 2004; Matthee *et al.* 2004; Nicolas *et al.* 2008; Tolley *et al.* 2008; Voelker *et al.* 2010). An alternative hypothesis suggests that forest refugia served as 'museums' preventing the loss of ancient diversity during the expansion of savannah areas (Fjeldså and Bowie 2008; Murienne *et al.* 2013). Both hypotheses have been invoked for African species, including avifauna, reptiles, mammals and invertebrates (Fjeldså and

Bowie 2008; Fjeldså and Lovett 1997; Matthee *et al.* 2004; Nicolas *et al.* 2008; Roy *et al.* 2001; Smitz *et al.* 2013; Tolley *et al.* 2008; Voelker *et al.* 2010), however, refugial diversity models have rarely been tested for African bats (but see Hassanin *et al.* 2015; Taylor *et al.* 2012).

Hipposideros is a highly speciose genus from the Old World family Hipposideridae, which exhibits a high level of cryptic diversity (Kingston *et al.* 2001; Thabah *et al.* 2006; Thong *et al.* 2012; Vallo *et al.* 2011; Vallo *et al.* 2008). The *Hipposideros caffer* complex, which occurs throughout Africa and in parts of the Arabian Peninsula, traditionally comprises two recognised, though cryptic, morphological species; Sundevall's roundleaf bat *Hipposideros caffer* (Sundevall, 1846) and Noack's roundleaf bat *Hipposideros ruber* (Noack, 1893). *Hipposideros caffer* is associated with wooded savannah, dry open and coastal forest habitats, while *H. ruber* occurs in rainforest and wet forested savannah (Monadjem *et al.* 2010; Vallo *et al.* 2008; Wright 2009). Several subspecies are recognised with substantial variability in morphology and it is suggested that further cryptic diversity may exist (Simmons *et al.* 2005; Vallo *et al.* 2008).

Mitochondrial (mtDNA) evidence has revealed the existence of four deeply divergent lineages within the *H. caffer* complex, which possibly constitute at least five species, rather than the two described (Vallo *et al.* 2008). Two smaller forms pertaining to the *H. caffer* morphotype, currently recognized as *H. caffer caffer* and *H. c. tephrus*, belong to mtDNA lineage A, with three further lineages, B, C, and D, broadly corresponding to the *H. ruber* morphotype (Vallo *et al.* 2008). However, mitochondrial data is limited by the nature of its maternal inheritance. Taxonomic classifications based on mtDNA can result in oversplitting (Elias *et al.* 2007; Lausen *et al.* 2008), as substantial nuclear gene flow can still occur even among highly divergent mitochondrial clades (Avise 2012; Castella *et al.* 2001; Petit and Mayer 1999). Taxonomic classifications must therefore be based on multiple lines of evidence (Cronin 1993; Mayer and von Helversen 2001). Microsatellites, which are bi-parentally inherited, are powerful tools to investigate gene flow, hybridisation or introgression between the mitochondrial lineages. Additional lines of data, such as morphometric and echolocation call frequency data, can be used

to further investigate divergence. In bats, which may be highly morphologically conserved, acoustic differences are often indicative of cryptic species (Jones and Barlow 2004). Such divergences may confer benefits related to species recognition, intraspecific communication, reproductive isolation, or resource partitioning (Jones and Barlow 2004; Kingston *et al.* 2001; Russo *et al.* 2007). Acoustic divergence may be more likely in species that employ high duty cycle echolocation, which comprise a dominant constant frequency component and have a highly sensitive acoustic fovea for frequency discrimination, such as bats of the Hipposideridae (Jones and Barlow 2004).

The taxonomic resolution of the *Hipposideros caffer* complex is of particular interest because they have been discovered to host coronaviruses (CoVs) with zoonotic potential (Chapter 3; Pfefferle *et al.* 2009), emphasizing the need for knowledge about their ecology. These viruses are related to several of high concern to human public health, including SARS-CoV and MERS-CoV, each responsible for recent outbreaks of severe respiratory disease with high case fatality (WHO 2004; Zaki *et al.* 2012), and share a putative evolutionary origin in bats, with other animal intermediate hosts (Haagmans *et al.* 2014; Ithete *et al.* 2013; Lau *et al.* 2005; Li *et al.* 2005). Contact between humans and bats, including *H. caffer* and *H. ruber*, is common in parts of Africa (Anti *et al.* 2015), and common practices such as bushmeat consumption and cave visitation have been directly linked to zoonotic outbreaks on the continent (Amman *et al.* 2012; Leroy *et al.* 2009; Swanepoel *et al.* 2007). There is therefore a pressing need for greater insight into the infection dynamics and ecology of potentially zoonotic CoVs in wildlife reservoirs, in order to inform prediction and prevention strategies where necessary.

This study aims to resolve some of the taxonomic uncertainty in bats in the *H. caffer* complex, and to investigate the processes that generate and maintain diversity in these bats. We use nuclear genetic markers (microsatellites), acoustic data, and morphological measurements to investigate cryptic speciation using representatives of all four major mitochondrial lineages. We look for evidence of hybridisation, and investigate differences in echolocation call frequency, and

consider explanations for acoustic divergence. We predict that previously reported mitochondrial divergences will be reflected in a lack of nuclear gene flow. Sympatrically roosting bat species may undergo acoustic divergence as a result of selective pressure for species recognition and intraspecific communication (Jones and Barlow 2004; Kingston *et al.* 2001). Given that the ranges of the four lineages overlap, and lineages B, C, and D often roost sympatrically (Chapter 6), we predict that an absence of gene flow between mtDNA lineages will be reflected in acoustic divergence. Using cytochrome *b* sequence data, we date past divergences, to examine hypotheses for species richness in African forests. Previous research has shown evidence of a major radiation in the *Hipposideros* genus approximately 15 – 14 million years ago (Ma), during the Mid-Miocene Climatic Optimum, therefore under the museum hypothesis we would expect the majority of divergences to occur around this period (Evans *et al.* 2004; Murienne *et al.* 2013), while the elevated speciation model predicts divergences to occur after the onset of climatic fluctuations in the late Miocene (Murienne *et al.* 2013).

MATERIALS AND METHODS

Sample collection

Samples were collected from nine bat colonies located in caves, abandoned mines or buildings in Ghana, West Africa (table 1, figure 1), where representatives of each of the four major lineages occur. Permission for capture was obtained from the Wildlife Division of the Forestry Commission of the Ministry of Lands, Forestry and Mines. Bats were captured with mist nets set at cave entrances or along trails between one hour after dusk and dawn. Bats were then sexed and forearm length and mass were measured using vernier calipers and an electronic balance. Juveniles were identified by incomplete fusion of the fourth epiphyseal joint (Brunet-Rossinni and Wilkinson 2009). Two 3mm wing biopsies were obtained using a biopsy punch for DNA extraction and stored in 90% ethanol.

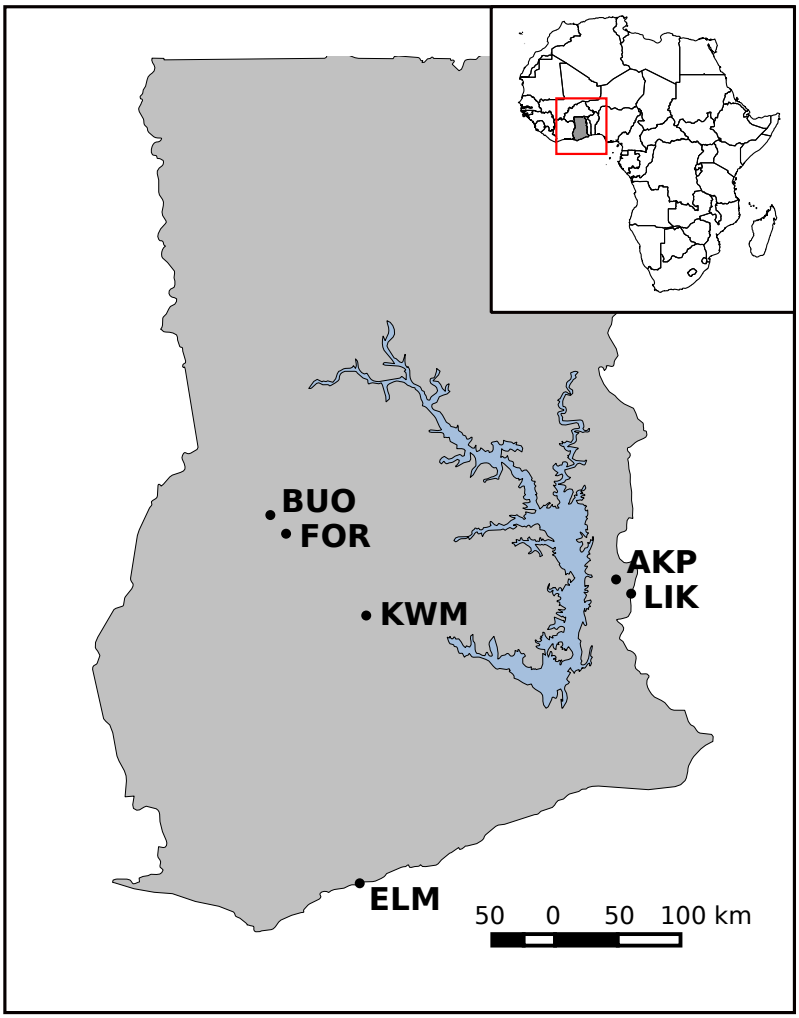


Figure 1. Sampling localities of *Hipposideros* cf. *ruber* used in this study. BUO = Buoyem, FOR = Forikrom, KWM = Kwamang, AKP = Akpafu-Todzi (hereafter "Akpafu"), LIK = Likpe-Todome (hereafter "Likpe"), ELM = Elmina.

DNA extraction and sequencing

Total genomic DNA was isolated from tissue samples using a QIAamp DNA micro kit (QIAGEN) or a GenCatch Genomic DNA Extraction kit (Epoch Life Science). Approximately 1140 bp of the cytochrome *b* gene was amplified by polymerase chain reaction (PCR) using the primers L14724 and H15915 (Irwin *et al.* 1991) or a slightly modified version L14724ag and H15915ag (Guillén-Servent and Francis 2006). Reactions were performed following Vallo *et al.* (2008) using *Taq* DNA polymerase (Promega). The PCR products were purified using a QIAquick PCR purification kit and sequenced commercially (Macrogen, Seoul, Korea), with the

same primers, using Big-Dye Terminator sequencing chemistry (Applied Biosystems, Foster City, CA, USA) on an ABI 3730xl sequencer. Sequences were aligned using BIOEDIT 7.1 (Hall 1999).

Table 1. Details of sampling localities of *Hipposideros* cf. *ruber* used in this study

Site	Site type	Latitude	Longitude
Akpafu-Todzi	mine	7.2619722	0.4915278
Buoyem 1	cave	7.7235833	-1.9879167
Buoyem 2	cave	7.7238056	-1.9926389
Elmina	building	5.0827778	-1.3483056
Forikrom	cave	7.58975	-1.8750833
Kwamang Cave 1	cave	7.0035685	-1.3003098
Kwamang Cave 2	cave	6.9832778	-1.2731944
Kwamang 3	cave	7.0065315	-1.3012354
Likpe-Todome 1	cave	7.1639444	0.6079167
Likpe-Todome 2	cave	7.1638611	0.6081389

Microsatellite genotyping

Nine microsatellite loci, isolated from *Hipposideros* aff. *ruber* (Hr2, Hr3, Hr5, Hr6, Hr7, Hr8, Hr10, Hr11, Hr12; Baldwin *et al.* 2014) were genotyped for 286 individuals. Primers were labelled with 5'-fluorochrome bases (VIC, FAM, NED or PET) and loci amplified following PCR protocols outlined in Baldwin *et al.* (2014). PCR products were electrophoresed using an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Allele sizes were determined via manual inspection using the software PEAK SCANNER 1.0 (Applied Biosystems), followed by automated binning performed using TANDEM 1.09. We reanalyzed 20% of individuals to ensure data integrity.

Genetic diversity and differentiation

Haplotype diversity (*h*), nucleotide diversity (π), and number of polymorphic sites were estimated for lineages B, C, and D from cyt *b* using DNASP 5.10 (Librado and Rozas 2009). Microsatellite markers were examined for deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium using GENEPOP 4.2 (Raymond and Rousset 1995). We analysed evidence for nuclear gene flow by performing a discriminant analysis of principal components (DAPC;

Jombart *et al.* 2010) and k-means cluster analysis on the microsatellite data using the ADEGENET package in R (R Core Team 2013). We looked for evidence of hybridization between the lineages by performing assignment tests in the program GENECLASS2 2.0 (Piry *et al.* 2004). For each individual, the likelihood of belonging to its identified mitochondrial lineage was calculated based on microsatellite genotype, with probability calculated by Monte Carlo Markov chain permutations. STRUCTURE 2.3 (Pritchard *et al.* 2000) was used to examine further evidence of hybridisation. Using the admixture model with independent allele frequencies, and lineage identities as POPID priors, ten replicate runs were performed with 80 000 MCMC iterations and a burn-in of 10 000 for k=2 to k=7.

Estimates of divergence time

To estimate divergence dates and times since the most recent common ancestors (TMRCA) among lineages, we used the Bayesian MCMC sampling method in the program BEAST 1.8.2 (Drummond and Rambaut 2007). Sequences of 1100 bp representative of the major clades of the four lineages from the *H. caffer* complex, several African congeneric species *H. beatus*, *H. fuliginosis*, *H. abae*, *H. gigas*, *H. jonesi* and *H. cyclops,* and several other bat species (*Nycteris thebaica, Coleura afra, Stenonycteris lanosus, Rousettus aegyptiacus, Rhinolophus landeri, Asellia tridens, Aselliscus stoliczkanus, Triaenops persiscus*) were included to improve the phylogenetic reconstruction and dating accuracy. Sequences not generated in this study were obtained from GenBank (Appendix 1). Sequences were analysed in two partitions, with the first comprising the first two codon positions, and the second comprising the third position, as determined by PARTITIONFINDER (Lanfear *et al.* 2012). The Hasegawa, Kishino and Yano model with gamin distribution and a proportion of invariant sites (HKY + G + I) was selected as the most appropriate model of sequence evolution for partition one and HKY + G for partition two by PARTITIONFINDER and JMODELTEST2 (Darriba *et al.* 2012) under Bayesian Information Criterion (BIC). An uncorrelated lognormal clock was used following comparison of marginal

likelihoods with strict, exponential and random local clock models using stepping-stone and path sampling (Baele *et al.* 2012; Baele *et al.* 2013).

Root age was set to 64 – 65 Ma following (Eick *et al.* 2005; Foley *et al.* 2015; Teeling *et al.* 2005). Node ages were calibrated using well-supported divergence dates based on molecular data reported by Teeling *et al.* (2005) and Foley *et al.* (2015). Specifically, these nodes were calibrated with normal priors and standard deviation corresponding to reported 95% credibility intervals, as follows: Yinpterochiroptera, 58 Ma; Rhinolophoidea, 51.5 Ma; *Hipposideros–Rhinolophus*, 40.5 Ma; *Hipposideros–Rhinonycteris*, 39 Ma; *Hipposideros–Asellia*, 34 Ma; *Hipposideros gigas/Aselliscus stoliczkanus*–monophyletic *Hipposideros* clade, 15 Ma; *H. jonesi*, 14 Ma; *H. abae*, 4 Ma. Tree models used a Yule process prior (Gernhard 2008). We ran four independent Monte Carlo Markov chains for 20 million generations each, sampling every 1000 generations, with 10% discarded as burn-in. Log files for the four chains were combined using LOGCOMBINER 1.8.2 and trees were summarised using TREEANNOTATOR 1.8.2 (Drummond and Rambaut 2007). Trees were visualised with FIGTREE 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Analysis of acoustic and morphometric variation

Echolocation calls of stationary (hand-held) bats were recorded using a handheld ultrasound detector (Pettersson D1000X; Pettersson Elektronik AB, Uppsala, Sweden) at a sampling rate of 400 kHz. Calls were analysed using MATLAB 8.0 (Mathworks Inc). To generate spectrograms, we used a fast fourier transformation (FFT) with a Hamming window of 1024 samples (2.6 ms) and an overlap of 98.4%. This resulted in a maximum frequency resolution of 390 Hz and a time resolution of 0.16 ms. The ten loudest calls per recording file were selected, each representing one individual. The FFT points of the constant frequency component, with an intensity between the maximum and 22 dB below, were extracted. Calls were selected to have a minimum call duration

of 4 ms to ensure that the constant frequency component was measured. The mean value of peak call frequency for each adult individual was used in the statistical analysis.

We compared differences in peak call frequency, and the morphological characteristics of forearm length and mass, using similar procedures. First, averages for adult bats for each lineage were compared for each character of interest using Kruskal-Wallis tests. Post-hoc Mann-Whitney-Wilcoxon tests were then performed for pairwise comparison. In order to account for the potential effects of other factors, generalised linear mixed models (GLMM) were also performed using the nlme package in R. For the GLMM on peak frequency, lineage, sex and forearm length were included as fixed variables. For forearm length, lineage and sex comprised the fixed effects. For the mass model, fixed effects were lineage and sex, with reproductive status as a random variable. All three models included site as a random variable.

RESULTS

Genetic diversity and differentiation

We sequenced the *cyt b* gene for 383 individuals, with 50 identified as lineage B, 66 lineage C, 263 lineage D and four lineage A (*H. caffer tephrus*). We found three lineages (B, C, D) occurring in sympatry at two sites (Buoyem), and lineages C and D occurred in sympatry at seven of the ten sites. Lineage D occurred at nine of ten sites, while the morphologically distinct form *H. caffer tephrus* (A) occurred allopatrically in one site (Elmina). Diversity statistics for each clade are shown in table 2.

There was no deviation from Hardy-Weinberg equilibrium for lineage A, (n = 12, H_O = 0.677, H_E = 0.663, F_{IS} = -0.023). There was evidence of null alleles in Hr1 and Hr13 for lineage B, Hr1 for lineage C and Hr13 for lineage D, and those loci were removed from analyses. For the remaining loci, no consistent departures from HWE or evidence of linkage disequilibrium was detected after Bonferroni correction (see table 2 in Chapter 6).

Table 2. Genetic variability at ten sites based on 782 bp of cyt *b*.

Lineage / Site	*n*	Haplotype diversity	Polymorphic sites	*d*	*h*	π
Lineage D (total)	263	23	37	6.893	0.772	0.0088
Akpafu	10	4	3	1.022	0.644	0.0013
Buoyem 1	11	4	3	0.691	0.491	0.0009
Buoyem 2	30	5	4	1.090	0.639	0.0014
Forikrom	35	5	5	0.521	0.429	0.0007
Kwamang 1	51	7	7	1.111	0.760	0.0014
Kwamang 2	48	3	2	0.232	0.228	0.0003
Kwamang 3	6	2	1	0.533	0.533	0.0007
Likpe 1	48	8	8	1.327	0.732	0.0017
Likpe 2	23	5	7	1.502	0.735	0.0019
Lineage C (total)	66	28	56	11.829	0.951	0.0151
Akpafu	27	13	28	6.336	0.920	0.0081
Buoyem 1	16	5	7	2.058	0.767	0.0026
Kwamang	7	5	7	2.857	0.905	0.0037
Likpe 2	15	27	50	11.704	0.952	0.0150
Lineage B (total)	50	6	8	0.358	0.227	0.0005
Buoyem 1	29	4	6	0.478	0.259	0.0006
Buoyem 2	19	3	2	0.211	0.205	0.0003
Lineage A (total)	4	1	1	0	0	0
Elmina	4	1	1	0	0	0

n = sample size, *d* = average number of pairwise nucleotide differences, *h* = haplotype diversity, π = nucleotide diversity

K-means clustering analysis identified four genetic clusters. The DAPC scatterplot shows separation of genetic groups (figure 2). Lineage D is clearly delineated from the other groups, as are B and C from each other. Lineage A shows slight overlap of 95% confidence ellipses with lineages B and C.

Assignment tests correctly assigned 283 of 286 individuals (quality index 90.3%). For the three individuals incorrectly assigned (one lineage A and two lineage D individuals assigned to lineage C), the probabilities of belonging to the incorrectly assigned lineages were low (0.12, 0.15, 0.05), and are therefore not strong evidence of any hybridization. STRUCTURE results showed that the number of clusters remained at a maximum of k=4, demonstrating clear delineation by lineage with high assignment probabilities and no strong evidence for hybridization (figure 3).

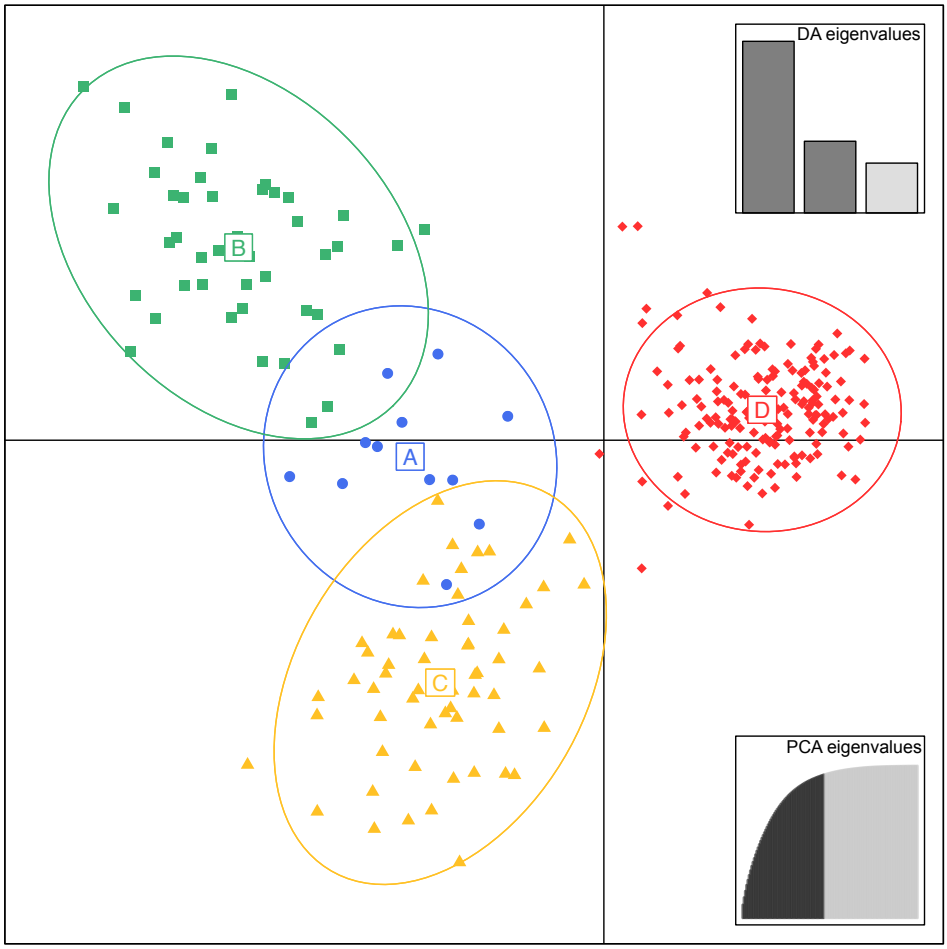


Figure 2. DAPC based on microsatellite markers for lineages A, B, C, and D.

Estimates of divergence time

Divergence time estimates generated in BEAST indicate that the African lineage of the *H. galeritus* group of hipposiderid species, to which the *H. caffer* complex, *H. fuliginosus, H. beatus,* and *H. abae* belong, diversified approximately 6.2 Ma, at the end of the Miocene (figure 4). *Hipposideros fuliginosus* is placed as the basal lineage in this group, with high bootstrap support. The consensus tree presents the *H. caffer* complex as a polyphyletic group, although bootstrap support for the placement of the nodes among recognised species and main *H. caffer* lineages was weak. This group, which includes the *H. caffer* lineages A, B, C, and D, *H. abae*, and *H. beatus*, is dated at 5.4 Ma, at the cusp of the Miocene-Pliocene epochs. Our analysis dated lineages A and C, and B and D, to approximately 4.5 Ma, although posterior probabilities between nominal

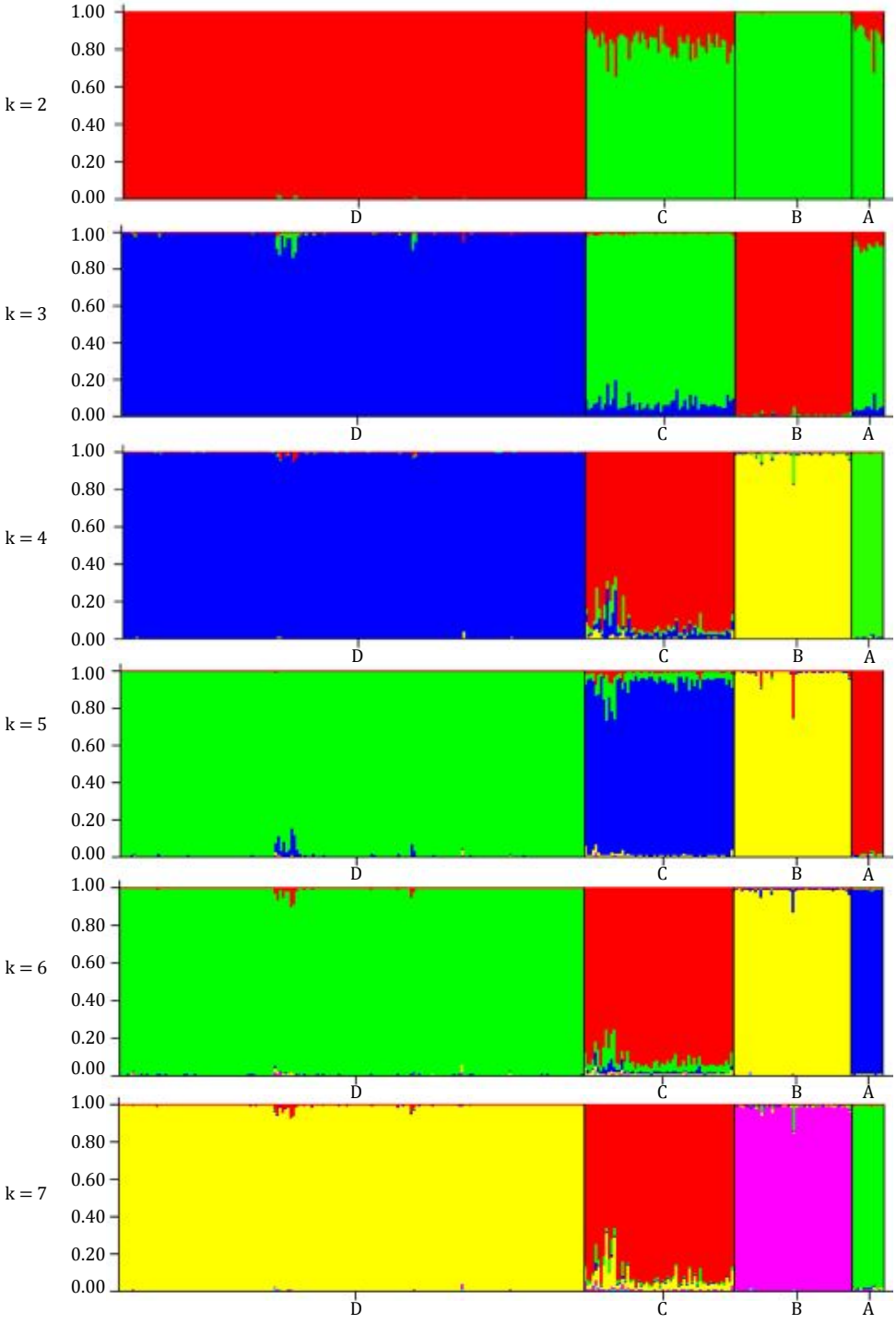


Figure 3. Structure results showing clustering of individuals belonging to mtDNA lineages D, C, B, and A, for k=2 to k=7.

sister clades are low. Divergences of sublineages within the major *H. caffer* complex clades appear have occurred between 1.0 and 3.6 Ma.

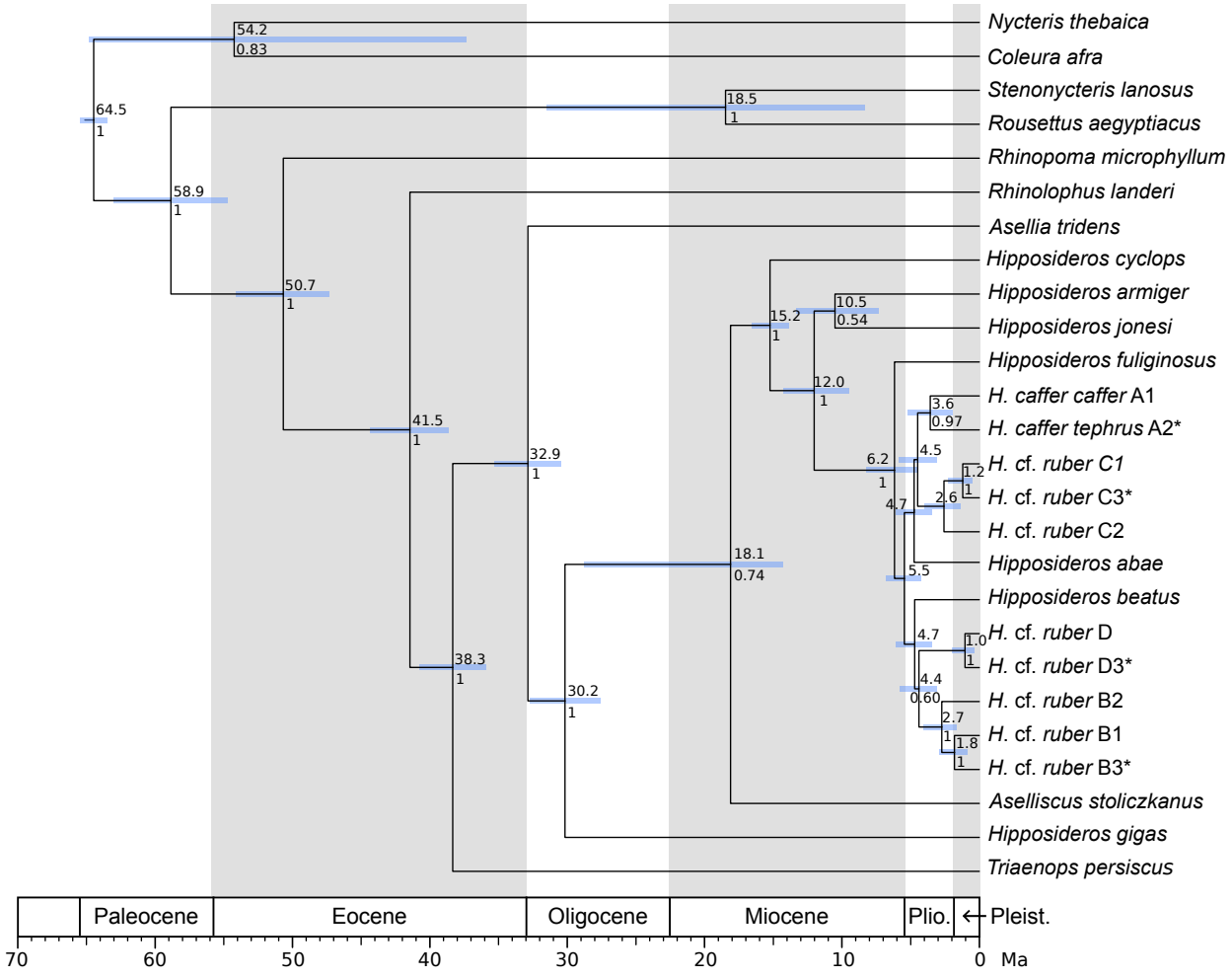


Figure 4. Maximum credibility tree based on cyt *b*. 95% credibility intervals for ages are represented by node bars. Posterior probabilities above 50% are indicated below nodes.

Acoustic and morphometric variation

Univariate tests and the GLMM showed highly significant differences in CF peak frequency, forearm length and mass between and among lineages (figure 5, tables 3 and 4). Echolocation call peak frequency differences ranged from 4.5 to 25 kHz for adjusted values (table 4), or 7 – 26.5 kHz for unadjusted values. Lineage A had the shortest forearm and the highest peak frequency,

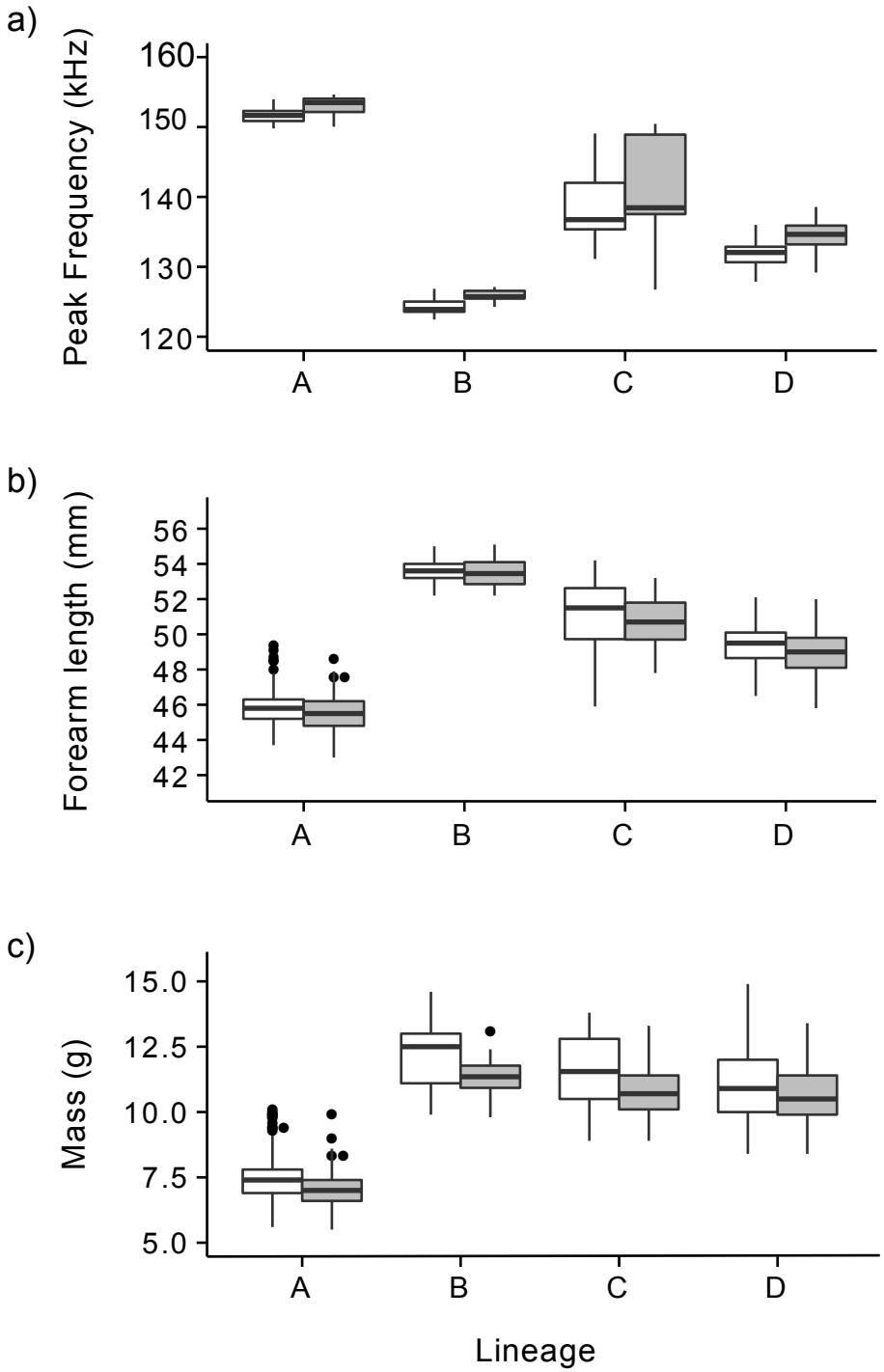


Figure 5. a) Mean call peak frequencies of adult bats, white = females, grey = males. Kruskal-Wallis χ^2 = 149.47, n = 202, df = 3, p < 0.00001; Bonferroni-adjusted Wilcoxon test p < 0.00001 for all pairwise comparisons; b) Forearm length by lineage of adult bats, white = females, grey = males. Kruskal-Wallis χ^2 = 766.02, n = 1342, df = 3, p < 0.00001; Bonferroni-adjusted Wilcoxon test p < 0.00001 for all pairwise comparisons; c) Mass by lineage of adult bats, white = females, grey = males. Kruskal Wallis χ^2 = 792.98, n = 1339, df = 3, p < 0.00001; Bonferroni-adjusted Wilcoxon test p < 0.00001 for A vs B, C and D, B vs D, p = 0.06 for B vs C, p = 0.11 for C vs D.

while lineage B had the largest forearm and the lowest call peak frequency. Lineage C exhibited a higher level of variation in call peak frequency and forearm length, with a difference in call peak frequency among sites (figure 5a,b). Call frequency at one site, Akpafu, was 10 – 15 kHz higher than at other sites (see Chapter 6). There were slight, but significant, sex differences in call peak frequency, forearm length and mass. The difference in mass between sexes was not significant when corrected for reproductive status (data not shown).

Sex was significantly associated with call peak frequency, forearm length and mass, with males calling at significantly higher frequencies and having significantly smaller forearm and mass measurements. However, these differences are likely smaller than our measurement resolution, and therefore may not have any biological meaning.

Table 3. Mean adult call peak frequency, forearm length and mass by lineage

Lineage	Peak Freq. (kHz)	Peak Freq. 95% CI	FA (mm)	FA 95%CI	Mass (g)	Mass 95% CI
A	151.9	148.6–155.2	45.7	43.6–47.8	7.3	5.8–8.8
B	125.4	122.6–128.1	53.5	51.8–55.2	11.8	9.6–14.1
C	140.2	127.9–152.5	50.9	47.2–54.5	11.2	8.4–14.0
D	133.2	128.3–138.0	49.1	46.4–51.9	10.8	8.2–13.4

Table 4. GLMM results for call peak frequency (n = 202, df = 187, R^2m = 0.80; R^2c = 0.94), forearm length (n = 1342, R^2m = 0.77; R^2c = 0.77, F-statistic 1135 on 4 and 1337 df) and mass (R^2m = 0.77; R^2c = 0.77, F-statistic 1111 on 4 and 1334 df).

Predictive variable	Call Peak Frequency		Forearm length		Mass	
	Mean Peak Frequency	SE (PF)	Mean FA length (mm)	SE (FA)	Mean Mass (g)	SE (Mass)
Intercept	156.87***	6.80	45.78***	0.04	7.61***	0.03
Lineage (ref = A)						
Lineage B	-24.73***	3.64	8.24***	0.17	5.01***	0.13
Lineage C	-13.62***	3.53	5.43***	0.15	4.05***	0.12
Lineage D	-18.11***	3.46	3.53***	0.08	3.68***	0.07
Sex (ref = F)	2.13***	0.32	-0.33***	0.07	-0.52***	0.06
FA	-0.12	0.13	—	—	—	—

***p < 0.0001

Standard deviation of the random intercept (residual) of random effects: peak frequency: site = 3.22 (2.16); forearm length: site = 0.55 (1.16); mass: site = 0.34 (0.58), reproductive status = 0.23 (0.48), residual = 0.71 (0.85).

DISCUSSION

We examined cryptic diversity in the *Hipposideros caffer* species complex by analysing genetic and acoustic divergence, and investigated speciation hypotheses by dating mtDNA divergences. We found representatives of all four major cyt *b* lineages described by Vallo *et al.* (2008) occurring in Ghana. Lineages B, C, and D often occurred in sympatric roosts in moist deciduous or transitional forest. The single colony of lineage A we identified in coastal savannah habitat was single-species, although they have also been found to roost sympatrically with the much larger *H. abae* (H. Baldwin, unpublished data). Interestingly, this lineage, which corresponds to *Hipposideros caffer tephrus*, previously observed only from Morocco to Senegal and an isolated population in Yemen (Vallo *et al.* 2008), and this study represents the first record of its occurrence in the wet tropics.

Cryptic species: molecular, acoustic and morphological divergence

We tested the hypothesis of Vallo *et al.* (2008) that *H. caffer* and *H. ruber* comprise multiple cryptic species by examining nuclear genetic, morphological, and acoustic divergences among four described mitochondrial lineages. Nuclear genetic evidence supports genetic isolation of each lineage, satisfying the genetic species concept (Baker and Bradley 2006). Reproductive isolation appears to be complete, satisfying the reproductive isolation requirement of the biological species concept (Dobzhansky and Dobzhansky 1937; Mayr 1942). Morphological differences in forearm size and mass provide phenetic support, although measurements overlap between some groups and therefore these morphological characters alone may not be diagnostic. Lineage A, with an affinity to the smaller *H. caffer* morphotype, can be discriminated with high confidence based on forearm and mass, and forearm length may assist in identification of lineages B, C, and D, particularly in discriminating B from D, as they show no overlap. Whether other morphological differences exist, for example in skull, wing, or limb measurements or dentition, requires further investigation. Significant acoustic divergences exist between all lineages, and are sufficient to

confidently discriminate lineages A and B from each other, and discriminate from C/D with reasonable confidence. The higher level of variation in acoustic and forearm measurements within lineage C, support previous suggestions of further cryptic diversity within this lineage (Vallo *et al.* 2011; Vallo *et al.* 2008).

Given the strength of molecular, acoustic, and morphological evidence, we support the proposal that at least four species should be recognised. However, the taxonomy of this group is highly complex, with deep divergences within the major lineages (Vallo *et al.* 2011; Vallo *et al.* 2008). Additional nuclear DNA, acoustic and morphological data are required to assess species delimitation within lineages, particularly within lineages A, B, and C, to illuminate the extent of cryptic diversity within the *H. caffer* group throughout the Afrotropics. In the interim, we propose provisional designation as *Hipposideros* species B, C, and D, and continued use of the name *H. caffer tephrus* for lineage A2, until further investigations can sufficiently elucidate the systematics of this group and assign appropriate nomenclature.

Diversification analysis

The *Hipposideros* genus is proposed to have radiated approximately 15 – 14 Ma, during the relatively warm and humid Mid-Miocene Climatic Optimum (Foley *et al.* 2015). Our data suggest that divergence in the African lineage of *H. galeritus* occurred much later, beginning in the late Miocene. The major *H. caffer* lineages, A, B, C, and D, *H. abae*, *H. beatus*, and *H. fuliginosus* diverged during the late Miocene and early Pliocene. Our results suggest that the *H. caffer* complex lineages may form a polyphyletic group, with A and C, and B and D, split into two groups with *H. abae* and *H. beatus*, respectively, although posterior probabilities were low. Additional molecular evidence, particularly from nuclear sequences, would be required to further comment on the phylogenetic status of the described *H. caffer* complex species in relation to others of the African *H. galeritus* group (*H. beatus*, *H. abae*, *H. fuliginosus*). In any case, the MRCAs of the main lineages within the monophyletic clade, comprising lineages A, B, C, D,

H. abae, and *H. beatus*, occurred between 5.5 and 4.4 million years ago, during the late Miocene/Early Pliocene. The common ancestors within extant representatives of the *H. caffer* complex dated to the late Pliocene and the Pleistocene (3.6 – 1 Ma). The new Ghanaian clades identified in this study (C3, D3, B3) are deeply diverged, with divergence dates between 2.5 and 1.3 Ma. These findings suggest fairly rapid diversification during the last six million years, and are consistent with the hypothesis that climate-driven forest dynamics since the late Miocene were instrumental in the diversification of this group. It has been observed that taxa that have diverged in the Pliocene or earlier tend to be reproductively isolated, consistent with our findings. Our results support the species pump hypothesis, with forest refugia promoting allopatric speciation, rather than serving as museums of ancient diversity.

From the late Miocene, global environmental shifts were characterised in the Afrotropics by a fluctuating trend of increased aridity and savannah expansion (deMenocal 1995; 2004; Jacobs 2004; Zachos *et al.* 2001). We observed diversification in the *H. galeritus* group consistent with reported patterns for African forests. The major lineages of the *H. caffer* complex, *H. abae* and *H. beatus* diverged between 4.7 and 4.4 Ma, consistent with a major diversification pulse in African rain forest trees (Couvreur *et al.* 2008). Divergences within lineages B, C, D, at approximately 2.7 Ma (B, C), 1.8 (B), and 1.0 Ma (B, C), correspond (within 0.2 Ma) to the three major intensified glacial periods and increased aridity since the late Pliocene (deMenocal 1995; 2004; Hamilton and Taylor 1992), key junctures in the evolution of other taxa (deMenocal 2004). The single dry savannah-adapted member of the *H. galeritus* group likely diverged from a moist forest- or rainforest-adapted common ancestor in the mid-Pliocene, possibly exploiting a new ecological niche in the expanding savannah of the period. Such enhanced speciation resulting from isolation in Miocene and Plio-Pleistocene forest refugia has been postulated as an important mechanism of speciation for other African plant and animal species (Couvreur *et al.* 2008; deMenocal 2004; Plana 2004), including birds (Bowie *et al.* 2004a; Bowie *et al.* 2004b; Fjeldså and Lovett 1997; Roy *et al.* 2001; Voelker *et al.* 2010), reptiles

(Daniels *et al.* 2007; Leaché and Fujita 2010; Tolley *et al.* 2006; Tolley *et al.* 2008), and mammals (Jacquet *et al.* 2014; Nicolas *et al.* 2008; Smitz *et al.* 2013).

Acoustic divergence: a mechanism of reproductive isolation?

In recently diverged, sympatric bat species, acoustic differences are thought to have evolved in allopatry, or as a result of allopatric speciation and secondary contact (Jones and Barlow 2004). Call frequency divergence may be non-adaptive, having arisen from drift or founder effects in isolated populations, with sufficient genetic or acoustic divergence by the time contact is re-established to prevent interbreeding (Jones and Barlow 2004). Alternatively, it may confer some selective advantage. It may arise by social selection, in order to facilitate recognition of conspecifics (Russo *et al.* 2007) or intraspecific communication, and avoid jamming or masking of calls from other species (Jones 1997; Jones and Barlow 2004; Kingston *et al.* 2001; Russo *et al.* 2007). It may be selected for in order to avoid hybridization (reproductive character displacement; Clare *et al.* 2013; Russo *et al.* 2007). Finally, it may serve to reduce interspecific competition by partitioning the sizes of targeted prey (resource partitioning; Heller and Helversen 1989; Jones and Barlow 2004; Puechmaille *et al.* 2011). These explanations are not mutually exclusive. Acoustic divergence that has arisen as a result of drift, for example, may diverge further upon secondary contact as a result of selective pressure. As relatively large differences in call frequency are required for functional differences in prey detection, we can expect that large differences in call frequency provide evidence for the resource partitioning hypothesis, while in the case of acoustic drift, social selection or character displacement, we would expect differences slight enough not to affect target detection but large enough to facilitate species recognition (Clare *et al.* 2013; Jones and Barlow 2004; Kingston *et al.* 2001).

We found significant variation in echolocation call frequencies between all major lineages in the *H. caffer* complex. The difference in call frequency is inversely correlated with significant differences in size (forearm length and mass) for the same groups, with the largest difference (25

kHz, A and B) reflecting the largest size difference between lineages. However, even this relatively large difference in call frequency is unlikely to have any bearing on prey recognition, corresponding to a difference in wavelength of only 0.48 mm. We propose that the acoustic divergence observed is likely a result of either drift (genetic or cultural) during previous isolation, social selection or reproductive character displacement to avoid hybridization. While it is likely not possible to disentangle the mechanisms of drift and social selection, comparing echolocation calls from non-overlapping parts of their ranges may help elucidate whether acoustic divergence has evolved as a form of reproductive character displacement, as areas of overlapping ranges should show the greatest divergence.

CONCLUSIONS

We provide multiple lines of evidence supporting the proposal that the major lineages of the *H. caffer* complex represent at least four distinct species. Further investigation from samples throughout the range is required to fully investigate potential cryptic diversity in its entirety within the *Hipposideros caffer* complex and to elucidate the complex systematics of the *H. galeritus* group.

Our results shed light on patterns and processes of diversity in African forest fauna. Our findings are consistent with the hypothesis that isolation in forest refugia promoted elevated speciation rates. Isolated, divergent populations likely expanded to currently observed distributions during warmer, wetter periods, leading to secondary contacts. This secondary contact may have promoted the observed acoustic divergence, either through drift, social selection, or reproductive character displacement. Further research would be required to test these theories.

The cryptic diversity described requires revision of conservation statuses and strategies to reveal any requirements for conservation priority or action. Given that these bats are important hosts of CoVs, it would be pertinent to know whether all of the cryptic species carry these

potentially zoonotic viruses. We have found CoVs in lineages A, D, and *H. abae*, however, the available individuals of *H.* sp. B and C were insufficient to conclude on their susceptibility (H. Baldwin, unpublished data). Further knowledge of the ecology of these species, including dispersal, population connectivity, and social behaviour would be valuable in predicting and preventing emerging infectious diseases in humans.

ACKNOWLEDGEMENTS

We are grateful to Isaac Mawusi Adanyeguh, Lucinda Kirkpatrick, Anna Vogeler, Mac Elikem Nutsuakor, David Ofori Agyei, Sarah Koschnicke, Julia Morrison, Emmanual Asare, Eunice Okyere, Kenneth Quansah, Emmanuel Essoun, Paul Marfo, Kennedy Darkwa, Stefan Klose, our guides Justice, Paul, Kwame Takye, Winfred, Samuel and Francis, and Thomas Kruppa for their help during the organization and conduct of field work. We thank the communities, community leaders and administrators of the sampling sites, the staff of the Ghana Wildlife Division, and the staff at Kumasi Centre for Collaborative Research in Tropical Medicine. Thank you to Paolo Momigliano and Lucinda Kirkpatrick for helpful discussion, and to Siobhan Dennison and Kate Umbers for comments on the manuscript. Sequences used in the phylogeny correspond to the following GenBank accession numbers: JN398215.1, DQ445714.1, EU934470.1, JF439015.1, EU434954.1, EU798757.1, KF874547.1, KJ735798.1, JQ710751.1, FJ347976.1, EU934468.1, FJ457612.1, EU934448.1.

REFERENCES

Amman B. R., Carroll S. A., Reed Z. D., Sealy T. K., Balinandi S., Swanepoel R., Kemp A., Erickson B. R., Comer J. A. & Campbell S. (2012) Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. *PLOS Pathogens* 8(10), e1002877.

Anti P., Owusu M., Agbenyega O., Annan A., Badu E. K., Nkrumah E. E., Oppong S. K., Adu-Sarkodie Y. & Drosten C. (2015) Human–bat interactions in rural West Africa. *Emerging Infectious Diseases* 21(8), 1418-1421.

Avise J. C. (2012) *Molecular markers, natural history and evolution*. Springer Science & Business Media, Sunderland, Massachusetts.

Baele G., Lemey P., Bedford T., Rambaut A., Suchard M. A. & Alekseyenko A. V. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 29(9), 2157-2167.

Baele G., Li W. L. S., Drummond A. J., Suchard M. A. & Lemey P. (2013) Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* 30(2), 239-243.

Baker R. J. & Bradley R. D. (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy* 87(4), 643-662.

Baldwin H., Vallo P., Gardner M., Drosten C., Tschapka M. & Stow A. (2014) Isolation and characterization of 11 novel microsatellite loci in a West African leaf-nosed bat, *Hipposideros* aff. *ruber*. *BMC Research Notes* 7(1), 607-610.

Bates P. J., Cameron K., Pearch M. J. & Hayes B. (2013) A review of the bats (Chiroptera) of the Republic of Congo, including eight species new to the country. *Acta Chiropterologica* 15(2), 313-340.

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

Bowie R. C., Fjeldså J., Hackett S. J. & Crowe T. M. (2004a) Molecular evolution in space and through time: mtDNA phylogeography of the olive sunbird (*Nectarinia olivacea*/*obscura*) throughout continental Africa. *Molecular Phylogenetics and Evolution* 33(1), 56-74.

Bowie R. C., Fjeldså J., Hackett S. J., Crowe T. M. & Fleischer R. (2004b) Systematics and biogeography of double-collared sunbirds from the Eastern Arc Mountains, Tanzania. *The Auk* 121(3), 660-681.

Brunet-Rossinni A. K. & Wilkinson G. S. (2009) Methods for age estimation and the study of senescence in bats. In: *Ecological and Behavioral Methods for the Study of Bats* (eds T. H. Kunz and S. Parsons) pp. 315-325. Johns Hopkins University Press, Baltimore.

Castella V., Ruedi M. & Excoffier L. (2001) Contrasted patterns of mitochondrial and nuclear structure among nursery colonies of the bat *Myotis myotis*. *Journal of Evolutionary Biology* 14(5), 708-720.

Ceballos G., Ehrlich P. R., Barnosky A. D., García A., Pringle R. M. & Palmer T. M. (2015) Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances* 1(5), e1400253.

Clare E. L., Adams A. M., Maya-Simões A. Z., Eger J. L., Hebert P. D. & Fenton M. B. (2013) Diversification and reproductive isolation: cryptic species in the only New World high-duty cycle bat, *Pteronotus parnellii*. *BMC Evolutionary Biology* 13(1), 26-43.

Couvreur T. L., Chatrou L. W., Sosef M. S. & Richardson J. E. (2008) Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biology* 6(1), 54-63.

Daniels S. R., Hofmeyr M. D., Henen B. T. & Crandall K. A. (2007) Living with the genetic signature of Miocene induced change: evidence from the phylogeographic structure of the endemic angulate tortoise *Chersina angulata*. *Molecular Phylogenetics and Evolution* 45(3), 915-926.

Darriba D., Taboada G. L., Doallo R. & Posada D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8), 772-772.

deMenocal P. B. (1995) Plio-Pleistocene African climate. *Science* 270(5233), 53-59.

deMenocal P. B. (2004) African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth and Planetary Science Letters* 220(1-2), 3-24.

Dobzhansky T. & Dobzhansky T. G. (1937) *Genetics and the Origin of Species*. Columbia University Press, New York.

Drummond A. & Rambaut A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7(1), 214-221.

Eick G. N., Jacobs D. S. & Matthee C. A. (2005) A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (Chiroptera). *Molecular Biology and Evolution* 22(9), 1869-1886.

Elias M., Hill R. I., Willmott K. R., Dasmahapatra K. K., Brower A. V., Mallet J. & Jiggins C. D. (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society of London B: Biological Sciences* 274(1627), 2881-2889.

Evans B. J., Kelley D. B., Tinsley R. C., Melnick D. J. & Cannatella D. C. (2004) A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular Phylogenetics and Evolution* 33(1), 197-213.

Fahr J. & Kalko E. K. (2011) Biome transitions as centres of diversity: habitat heterogeneity and diversity patterns of West African bat assemblages across spatial scales. *Ecography* 34(2), 177-195.

Fjeldså J. & Bowie R. C. (2008) New perspectives on the origin and diversification of Africa's forest avifauna. *African Journal of Ecology* 46(3), 235-247.

Fjeldså J. & Lovett J. C. (1997) Geographical patterns of old and young species in African forest biota: the significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation* 6(3), 325-346.

Foley N. M., Thong V. D., Soisook P., Goodman S. M., Armstrong K. N., Jacobs D. S., Puechmaille S. J. & Teeling E. C. (2015) How and why overcome the impediments to

resolution: lessons from rhinolophid and hipposiderid bats. *Molecular Biology and Evolution* 32(2), 313-333.

Furman A., Postawa T., Öztunç T. & Çoraman E. (2010) Cryptic diversity of the bent-wing bat, *Miniopterus schreibersii* (Chiroptera: Vespertilionidae), in Asia Minor. *BMC Evolutionary Biology* 10(1), 121-132.

Gernhard T. (2008) The conditioned reconstructed process. *Journal of Theoretical Biology* 253(4), 769-778.

Guillén-Servent A. & Francis C. M. (2006) A new species of bat of the *Hipposideros bicolor* group (Chiroptera: Hipposideridae) from central Laos, with evidence of convergent evolution with Sundaic taxa. *Acta Chiropterologica* 8(1), 39-61.

Haagmans B. L., Al Dhahiry S. H. S., Reusken C. B. E. M., Raj V. S., Galiano M., Myers R., Godeke G.-J., Jonges M., Farag E., Diab A., Ghobashy H., Alhajri F., Al-Thani M., Al-Marri S. A., Al Romaihi H. E., Al Khal A., Bermingham A., Osterhaus A. D. M. E., AlHajri M. M. & Koopmans M. P. G. (2014) Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *The Lancet Infectious Diseases* 14(2), 140-145.

Hall T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic Acids Symposium Series* pp. 95-98.

Hamilton A. & Taylor D. (1992) History of climate and forests in tropical Africa during the last 8 million years. *Tropical Forests and Climate* 19, 65-78.

Hassanin A., Khouider S., Gembu G.-C., Goodman S. M., Kadjo B., Nesi N., Pourrut X., Nakouné E. & Bonillo C. (2015) The comparative phylogeography of fruit bats of the tribe Scotonycterini (Chiroptera, Pteropodidae) reveals cryptic species diversity related to African Pleistocene forest refugia. *Comptes Rendus Biologies* 338(3), 197-211.

Heller K. G. & Helversen O. v. (1989) Resource partitioning of sonar frequency bands in rhinolophoid bats. *Oecologia* 80, 178-186.

Irwin D. M., Kocher T. D. & Wilson A. C. (1991) Evolution of the cytochrome *b* gene of mammals. *Journal of Molecular Evolution* 32(2), 128-144.

Ithete N. L., Stoffberg S., Corman V. M., Cottontail V. M., Richards L. R., Schoeman M. C., Drosten C., Drexler J. F. & Preiser W. (2013) Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerging Infectious Diseases* 19(10), 1697-1699.

Jacobs B. F. (2004) Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359(1450), 1573-1583.

Jacquet F., Nicolas V., Colyn M., Kadjo B., Hutterer R., Decher J., Akpatou B., Cruaud C. & Denys C. (2014) Forest refugia and riverine barriers promote diversification in the West African pygmy shrew (*Crocidura obscurior* complex, Soricomorpha). *Zoologica Scripta* 43(2), 131-148.

Jombart T., Devillard S. & Balloux F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11(1), 94.

Jones G. (1997) Acoustic signals and speciation: the roles of natural and sexual selection in the evolution of cryptic species. *Advances in the Study of Behaviour* 26, 317-354.

Jones G. & Barlow K. (2004) Cryptic species of echolocating bats. In: *Echolocation in bats and dolphins* (eds J. A. Thomas, C. F. Moss and M. Vater) pp. 345-349. University of Chicago Press, London.

Kingston T., Lara M., Jones G., Akbar Z., Kunz T. & Schneider C. (2001) Acoustic divergence in two cryptic *Hipposideros* species: a role for social selection? *Proceedings of the Royal Society Biological Sciences Series B* 268(1474), 1381-1386.

Lanfear R., Calcott B., Ho S. Y. & Guindon S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29(6), 1695-1701.

Lau S., Woo P., Li K., Huang Y., Tsoi H., Wong B., Wong S., Leung S., Chan K. & Yuen K. (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences of the United States of America* 102(39), 14040.

Lausen C., Delisle I., Barclay R. & Strobeck C. (2008) Beyond mtDNA: nuclear gene flow suggests taxonomic oversplitting in the little brown bat (*Myotis lucifugus*). *Canadian Journal of Zoology* 86(7), 700-713.

Leaché A. D. & Fujita M. K. (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B: Biological Sciences* 277(1697), 3071-3077.

Leroy E. M., Epelboin A., Mondonge V., Pourrut X., Gonzalez J.-P., Muyembe-Tamfum J.-J. & Formenty P. (2009) Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector-Borne and Zoonotic Diseases* 9(6), 723-728.

Li W., Shi Z., Yu M., Ren W., Smith C., Epstein J., Wang H., Crameri G., Hu Z., Zhang H., Zhang J., McEachern J., Field H., Daszak P., Eaton B. T., Zhang S. & Wang L. F. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310(5748), 676-679.

Librado P. & Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11), 1451-1452.

Mace G. M. & Purvis A. (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Molecular Ecology* 17(1), 9-19.

Maley J. (1989) Late Quaternary climatic changes in the African rain forest: forest refugia and the major role of sea surface temperature variations. In: *Paleoclimatology and Paleometeorology: Modern and Past Patterns of Global Atmospheric Transport* pp. 585-616. Springer, London.

Malhi Y., Adu-Bredu S., Asare R. A., Lewis S. L. & Mayaux P. (2013) African rainforests: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1625), 20120312.

Matthee C. A., Tilbury C. R. & Townsend T. (2004) A phylogenetic review of the African leaf chameleons: genus *Rhampholeon* (Chamaeleonidae): the role of vicariance and climate change in speciation. *Proceedings of the Royal Society of London B: Biological Sciences* 271(1551), 1967-1975.

Mayr E. (1942) *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Harvard University Press, New York.

Monadjem A., Richards L., Taylor P. J., Denys C., Dower A. & Stoffberg S. (2013a) Diversity of Hipposideridae in the Mount Nimba massif, West Africa, and the taxonomic status of *Hipposideros lamottei*. *Acta Chiropterologica* 15(2), 341-352.

Monadjem A., Richards L., Taylor P. J. & Stoffberg S. (2013b) High diversity of pipistrelloid bats (Vespertilionidae: *Hypsugo*, *Neoromicia*, and *Pipistrellus*) in a West African rainforest with the description of a new species. *Zoological Journal of the Linnean Society* 167(1), 191-207.

Monadjem A., Taylor P. J., Cotterill W. & Schoeman M. (2010) *Bats of Southern and Central Africa: A Biogeographic and Taxonomic Synthesis*. Wits University Press, Johannesburg.

Moritz C. (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology* 51(2), 238-254.

Moritz C. & Faith D. P. (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* 7(4), 419-429.

Murienne J., Benavides L. R., Prendini L., Hormiga G. & Giribet G. (2013) Forest refugia in Western and Central Africa as 'museums' of Mesozoic biodiversity. *Biology Letters* 9(1), 20120932.

Myers N., Mittermeier R. A., Mittermeier C. G., da Fonseca G. A. B. & Kent J. (2000) Biodiversity hotspots for conservation priorities. *Nature* 403(6772), 853-858.

Nicolas V., Mboumba J. F., Verheyen E., Denys C., Lecompte E., Olayemi A., Missoup A. D., Katuala P. & Colyn M. (2008) Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. *Journal of Biogeography* 35(11), 2074-2089.

Petit E. & Mayer F. (1999) Male dispersal in the noctule bat (*Nyctalus noctula*): where are the limits? *Proceedings of the Royal Society of London B: Biological Sciences* 266(1430), 1717-1722.

Pfefferle S., Oppong S., Drexler J., Gloza-Rausch F., Ipsen A., Seebens A., Müller M., Annan A., Vallo P., Adu-Sarkodie Y., Kruppa T. & Drosten C. (2009) Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerging Infectious Diseases* 15(9), 1377-1384.

Piry S., Alapetite A., Cornuet J.-M., Paetkau D., Baudouin L. & Estoup A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95(6), 536-539.

Plana V. (2004) Mechanisms and tempo of evolution in the African Guineo–Congolian rainforest. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359(1450), 1585-1594.

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

Puechmaille S. J., Gouilh M. A., Piyapan P., Yokubol M., Mie K. M., Bates P. J., Satasook C., Nwe T., Bu S. S. H. & Mackie I. J. (2011) The evolution of sensory divergence in the context of limited gene flow in the bumblebee bat. *Nature Communications* 2, 573.

R Core Team. (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raymond M. & Rousset F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86(3), 248-249.

Rouget M., Cowling R. M., Pressey R. L. & Richardson D. M. (2003) Identifying spatial components of ecological and evolutionary processes for regional conservation planning in the Cape Floristic Region, South Africa. *Diversity and Distributions* 9(3), 191-210.

Roy M. S., Sponer R. & Fjeldså J. (2001) Molecular systematics and evolutionary history of akalats (genus *Sheppardia*): a pre-Pleistocene radiation in a group of African forest birds. *Molecular Phylogenetics and Evolution* 18(1), 74-83.

Russo D. & Jones G. (2000) The two cryptic species of *Pipistrellus pipistrellus* (Chiroptera: Vespertilionidae) occur in Italy: evidence from echolocation and social calls. *Mammalia* 64(2), 187-198.

Russo D., Mucedda M., Bello M., Biscardi S., Pidinchedda E. & Jones G. (2007) Divergent echolocation call frequencies in insular rhinolophids (Chiroptera): a case of character displacement? *Journal of Biogeography* 34(12), 2129-2138.

Simmons N. B., Wilson D. & Reeder D. (2005) Order Chiroptera. In: *Mammal species of the world: a taxonomic and geographic reference* (eds D. E. Wilson and D. M. Reeder) pp. 312-529. Johns Hopkins University Press, Washington DC.

Smitz N., Berthouly C., Cornélis D., Heller R., Van Hooft P., Chardonnet P., Caron A., Prins H., van Vuuren B. J., De Iongh H. & Michaux J. (2013) Pan-African genetic structure in the African buffalo (*Syncerus caffer*): investigating intraspecific divergence. *PLOS One* 8(2), e56235.

Swanepoel R., Smit S., Rollin P., Formenty P., Leman P., Kemp A., Burt F., Grobbelaar A., Croft J. & Bausch D. (2007) Studies of reservoir hosts for Marburg virus. *Emerging Infectious Diseases* 13(12), 1847-1851.

Taylor P. J., Stoffberg S., Monadjem A., Schoeman M. C., Bayliss J. & Cotterill F. P. (2012) Four new bat species (*Rhinolophus hildebrandtii* complex) reflect Plio-Pleistocene divergence of dwarfs and giants across an Afromontane archipelago. *PLOS One* 7(9), e41744.

Teeling E. C., Springer M. S., Madsen O., Bates P., O'Brien S. J. & Murphy W. J. (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307(5709), 580-584.

Thabah A., Rossiter S. J., Kingston T., Zhang S., Parsons S., Mya K. M., Akbar Z. & Jones G. (2006) Genetic divergence and echolocation call frequency in cryptic species of *Hipposideros larvatus s.l.* (Chiroptera: Hipposideridae) from the Indo - Malayan region. *Biological Journal of the Linnean Society* 88(1), 119-130.

Thong V. D., Puechmaille S. J., Denzinger A., Bates P. J., Dietz C., Csorba G., Soisook P., Teeling E. C., Matsumura S. & Furey N. M. (2012) Systematics of the *Hipposideros turpis* complex and a description of a new subspecies from Vietnam. *Mammal Review* 42(2), 166-192.

Tolley K. A., Burger M., Turner A. A. & Matthee C. A. (2006) Biogeographic patterns and phylogeography of dwarf chameleons (*Bradypodion*) in an African biodiversity hotspot. *Molecular Ecology* 15(3), 781-793.

Tolley K. A., Chase B. M. & Forest F. (2008) Speciation and radiations track climate transitions since the Miocene Climatic Optimum: a case study of southern African chameleons. *Journal of Biogeography* 35(8), 1402-1414.

Vallo P., Benda P., Martínková N., Kaňuch P., Kalko E. K., Červený J. & Koubek P. (2011) Morphologically uniform bats *Hipposideros* aff. *ruber* (Hipposideridae) exhibit high

mitochondrial genetic diversity in southeastern Senegal. *Acta Chiropterologica* 13(1), 79-88.

Vallo P., Guillén-Servent A., Benda P., Pires D. & Koubek P. (2008) Variation of mitochondrial DNA in the *Hipposideros caffer* complex (Chiroptera: Hipposideridae) and its taxonomic implications. *Acta Chiropterologica* 10(2), 193-206.

Voelker G., Outlaw R. K. & Bowie R. C. (2010) Pliocene forest dynamics as a primary driver of African bird speciation. *Global Ecology and Biogeography* 19(1), 111-121.

World Health Organization. (2004) Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. http://www.who.int/csr/sars/country/table2004_2004_2021/en/.

Wright G. (2009) *Hipposideros caffer* (Chiroptera: Hipposideridae). *Mammalian Species* 845, 1-9.

Zachos J., Pagani M., Sloan L., Thomas E. & Billups K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292(5517), 686-693.

Zaki A. M., van Boheemen S., Bestebroer T. M., Osterhaus A. D. M. E. & Fouchier R. A. M. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine* 367, 1814-1820.

SUPPLEMENTARY MATERIAL

Figure S1. Photographs of *Hipposideros caffer tephrus* and *H.* sp. B, C, and D, in different colour phases. a) to d) *H. caffer tephrus*; e) to h) *H.* sp. B; i) to l) *H.* sp. C; m) to p) *H.* sp. D.

CHAPTER 6
Do sympatric sister species with similar traits show similar patterns of genetic variation? A case study in Afrotropical bats (Hipposideridae: *Hipposideros*)

Image: Marco Tschapka

H. J. Baldwin, P. Vallo, A. T. Ruiz, P. Anti, E. E. Nkrumah, E. K. Badu, E. K. V. Kalko, M. Tschapka, A. J. Stow.

Formatted for *Molecular Ecology*

ABSTRACT

Species that share distributions, life history and ecological traits are often assumed to have similar responses to historical events and dispersal barriers, and thus concordant spatial and temporal patterns of genetic structure. Here we examine patterns of intraspecific genetic variation in three closely related, sympatric bat species in the genus *Hipposideros* with similar ecological and natural history traits. Despite this high trait similarity and close relatedness, these species show stark contrasts in population genetic patterns, with marked differences in phylogeography and gene flow, and demographic history. *Hipposideros* sp. B exhibited very low genetic diversity, with evidence of a recent reduction in effective population size, while *H.* sp. C showed high diversity and a large, constant historical population size. *Hipposideros* sp. D showed evidence of a recent population expansion in western Ghana, and constant population size in the eastern sites. Steady isolation by distance for *H.* sp. D contrasted with localised movement patterns for *H.* sp. C, and no differentiation between sites for *H.* sp. B, albeit within a smaller range. These findings highlight the need to be wary of inferring intraspecific patterns of genetic variation based on phylogenetic and trait similarity.

Keywords: *Hipposideros caffer*, *Hipposideros ruber,* Hipposideridae, cytochrome *b*, microsatellites, demographic history, comparative phylogeography, gene flow, genetic connectivity

INTRODUCTION

Patterns of intraspecific genetic variation and demography are shaped by an organism's response to extrinsic geographical and climatic processes. Species-specific characteristics such as environmental requirements, life history traits, dispersal abilities, and ecological associations may influence these responses (Dawson 2014; Gutiérrez-García and Vázquez-Domínguez 2011). It is often assumed that species that share distributions, life history and ecological traits should show similar responses to historical events and dispersal barriers, and thus concordant spatial and

temporal patterns of genetic structure (Dawson 2012). This assumption can underpin decision-making in wildlife management, through extrapolation of knowledge about one species on the basis of phylogenetic relatedness or trait similarity to redress a lack of information on the target species. This approach, a consequence of limited resources and the urgency of conservation management, may be used to inform decisions such as identifying areas of conservation priority, dispersal corridors and barriers, and assessing risks posed by pathogens, parasites, and environmental pollution (e.g. Chittick *et al.* 2001; Sappington *et al.* 2001; Warman *et al.* 2004; Weir *et al.* 2010).

Intraspecific genetic diversity confers the evolutionary potential for resilience to changes in environment, disease outbreaks and other catastrophes, can increase reproductive viability, and may lead to speciation (Mace and Purvis 2008; Moritz and Faith 1998). While the focus of conservation is often on diversity patterns, preservation of evolutionary processes is essential for long-term population persistence, particularly in the face of the continued threats to biodiversity (Mace and Purvis 2008; Moritz 2002; Moritz and Faith 1998; Rouget *et al.* 2003). This requires specific planning based on knowledge of these processes that underpin evolutionary change in natural populations.

Comparative phylogeography provides a powerful analytical framework for revealing historical processes that shape genetic diversity within co-distributed species (Avise 2009). Such studies have elucidated the impacts of particular climatic and orographic events, and existing geographical features such as mountains and water bodies, and anthropogenic habitat fragmentation, which can act as dispersal barriers and drive gene flow, genetic divergence and speciation. Similarly, comparing intraspecific patterns of gene flow and genetic structure among co-distributed species can help to elucidate common landscape-level barriers to gene flow. Comparative approaches to investigating genetic patterns in related, co-distributed taxa known to differ in a specific trait, are also used to draw conclusions about the effects of specific traits in species responses to their environments (e.g. Dawson 2014; Paz *et al.* 2015). However, the

fundamental assumption that phylogenetically related organisms have similar responses to similar processes has not been well-examined as a result of the complexities of observational data (Dawson 2012; 2014). 'Natural experiments' set up by co-distributed sister taxa with similar life histories and ecologies provide an excellent opportunity to test the null hypothesis with relative control for confounding variables and noise (Dawson 2014).

Bats in the *Hipposideros caffer* group provide a fitting model for examining whether spatial and temporal patterns of genetic structure differ in co-occurring taxa with similar life history and ecological traits. We focus on three recently identified cryptic species within this Afrotropical group, traditionally typed as *Hipposideros ruber* (referred to as *H.* sp. B, C and D; Chapter 5), in Ghana, West Africa, where their ranges overlap (Chapter 5; Vallo *et al.* 2008). Divergence of these taxa dates to the late Pliocene, and distinct intraspecific mtDNA subclades arose in each during the early Pleistocene (Chapter 5). These bats are highly morphologically and ecologically similar, to the extent that diversity representing at least three distinct species has remained cryptic until recently (Chapter 5; Vallo *et al.* 2008). They have similar habitat preferences, often sharing roosts, have concurrent reproductive seasons, and similar diet and dispersal abilities, which are limited by conserved wing morphology for cluttered space (table 1; Chapter 5; Bell and Fenton 1983; Bell and Fenton 1984; Dunning and Krüger 1996; Norberg and Rayner 1987). This group therefore provides a natural experiment by which to test whether co-distributed taxa with similar life histories and ecological traits will have the expected highly concordant phylogeographic structure (Dawson 2012).

Further, these bats represent an interesting group in which to examine patterns of genetic diversity because they are known to harbor coronaviruses (CoVs) related to several viruses of public health concern (Chapter 3; Appendix 2; Pfefferle *et al.* 2009). Patterns of genetic diversity and dispersal, influenced by landscape heterogeneity or social and behavioural factors, can contribute to disease establishment and determine patterns of pathogen persistence and transmission (Archie *et al.* 2009; Biek and Real 2010; Guivier *et al.* 2011; Real and Biek 2007).

Intraspecific diversity may greatly contribute to buffering populations against epidemics (for review see Altizer *et al.* 2003). Knowledge of their spatial and dispersal dynamics may therefore be important for making epidemiological predictions, assessing risks for disease spillover and spread, and for the management and prevention of disease outbreaks.

This study aims to characterize patterns of genetic diversity within three *Hipposideros* species in Ghana, West Africa. We use comparative phylogeographic, population genetic, and coalescent methods to test for congruence in patterns of genetic structuring and demographic history in these closely related bat species. By utilising nuclear genotypic, allele frequency, and mitochondrial sequence data, we address these questions at both ecological and evolutionary time scales. As acoustic divergence can indicate population isolation and the effects of drift (Jones and Barlow 2004), we investigated intraspecific acoustic divergences among sites to further explore geographic differences. Exploring genetic diversity where the ranges of these sister species overlap provides an opportunity to assess whether co-distributed species with similar life history and ecological traits have the same responses to historical and contemporary processes and geographical features across space and time.

MATERIALS AND METHODS

Sampling, DNA extraction and amplification

Samples were collected from nine bat colonies located in natural caves, abandoned mines and buildings in the Ashanti, Brong Ahafo and Volta regions of central and eastern Ghana (figure 1b; for further details see Chapter 5). The bat colonies are located in upper Guinean rainforest (Kwamang, Buoyem, Forikrom sites), and the Togo Hills, a rainforest 'island' separated by dry savannah of the Dahomey Gap (Akpafu and Likpe sites). DNA was extracted and the cytochrome *b* gene amplified for 379 individuals and sequenced as described previously (Chapter 5), with sequences trimmed to 782 bp in order to maximise the included number of sequences. Bats were genotyped at 12 microsatellite loci (Hr1, Hr2, Hr3, Hr5, Hr6, Hr7, Hr8, Hr9, Hr10, Hr11, Hr12, Hr13; Baldwin *et al.* 2014) for 272 individuals. Primers were labelled with 5'-

fluorochrome bases (VIC, FAM, NED or PET) and PCRs protocols and conditions followed Baldwin *et al.* (2014). PCR products were electrophoresed using an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Allele sizes were determined via manual inspection using the software PEAK SCANNER 1.0 (Applied Biosystems), followed by automated binning performed using TANDEM 1.09. We reanalysed 20% of individuals for data integrity.

Table 1. Ecological characteristics of *Hipposideros* sp. B, C, D (*Hipposideros* cf. *ruber*), based on capture data from Ghana (this study), and Vallo *et al.* (2008)

Traits	**Species**			**References**
	***H.* sp. B**	***H.* sp. C**	***H.* sp. D**	
FA (95% CI) [mm]	53.5 (51.8–55.2)	50.9 (47.2–54.5)	49.1 (46.4–51.9)	Chapter 5
Mass [g]	11.8 (9.6–14.1)	11.2 (8.4–14.0)	10.8 (8.2–13.4)	Chapter 5
Habitat	rainforest/moist forest	rainforest/moist forest	rainforest/moist forest	Chapter 5, this study; Vallo 2008
Known range	West Africa (Upper Guinean forests; lineage B2), Coastal East Africa (lineage B1)	Western and Central Africa (Guineo-Congolian forests)	West Africa (upper Guinean forests)	Vallo 2008; this study
Reproductive season	Birth and lactation period March - May	Birth and lactation period March - May	Birth and lactation period March - May	this thesis

FA = forearm length. Habitat and reproductive data based on colonies and bats captured in this study (B: 3 sites, n = 50; C: 8 sites, n = 66; D: 7 sites, n = 263)

Analytical time scales

Investigations of genetic diversity at different temporal scales require subtly different analytical approaches. We employ a number of genic and genotypic approaches that analyse genetic diversity at different time scales. Methods based on neutral allele frequencies and genetic distance, such as Analysis of Molecular Variance (AMOVA), fixation indices (F_{ST}, G''_{ST}), and discriminant analysis of principal components (DAPC), reveal gradual process of genetic drift. On the other hand, spatial autocorrelation measures the shuffling of genotypic arrays, allowing variation to accumulate rapidly through recombination in each generation, providing insight into contemporary processes.

Phylogeographic analyses

To examine phylogenetic relationships within lineages, we employed the Bayesian MCMC sampling method in the program BEAST 1.8.2 (Drummond and Rambaut 2007). All unique cyt *b* haplotypes sampled, in addition to representative sequences from major intraspecific sublineages (of *H.* spp. A, B, C, and D), were included (Chapter 5), with *Rhinolophus landeri* as outgroup.

Sequences were analysed as a single partition following analysis in the program PARTITIONFINDER (Lanfear *et al.* 2012). The Hasegawa, Kishino and Yano model with gamma distribution (HKY + G) was selected as the most appropriate model of sequence evolution, as determined by JMODELTEST2 (Darriba *et al.* 2012) under Bayesian Information Criterion (BIC). We ran four independent Monte Carlo Markov chains for 20 million generations each, sampling every 1000 generations, with 10% discarded as burn-in. Log files for the four chains were combined using LOGCOMBINER 1.8.2 and trees were summarised using TREEANNOTATOR 1.8.2 (Drummond and Rambaut 2007). The consensus tree was visualised with FIGTREE 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). K2P pairwise genetic distances were calculated in MEGA 6.6 (Tamura *et al.* 2013).

Demographic history analyses

We inferred historical demography for each species using mtDNA diversity statistics, mismatch distribution analysis and neutrality tests, calculated in DNASP. High values of diversity statistics *h* and π indicate constant, large historical population sizes, while a high *h* and low π suggest recent population expansion (Russell *et al.* 2005). Similarly, a larger expansion coefficient (*S*/*d*; (Peck and Congdon 2004)) can infer historic population growth. We calculated the raggedness statistic (*rg*) from mismatch distribution, as well as the more powerful Fu's F_S, and Fu and Li's F^* and D^* (Fu 1997; Fu and Li 1993). Significance of *rg* and Fu's F_S were calculated using 1000

bootstrap replicates. Minimum-spanning networks were constructed in the program POPART (http://popart.otago.ac.nz).

Where population expansion was observed in these analyses, we used the program BOTTLENECK 1.2 which uses microsatellite allele frequencies to detect recent effective population size reductions, and therefore can be used to differentiate founder effects from bottlenecks (Cornuet and Luikart 1996). Because allele diversity is reduced faster than heterozygosity, an excess of heterozygosity relative to that predicted under mutation-drift equilibrium is expected for populations that have experienced a recent reduction (Cornuet and Luikart 1996). We used a sign test and a Wilcoxon signed rank test to test the significance of relative heterozygote excess under three models of evolution; the infinite allele model, step-wise mutation model, and a two-phase model that allows multiple-step mutations. We also tested for deviation from an L-shaped allele frequency distribution expected under mutation-drift equilibrium, as recent bottlenecks may provoke a mode shift.

Population genetic analyses

For the microsatellite loci, observed (H_O) and expected (H_E) heterozygosities and allelic richness (AR) were calculated using the DIVERSITY package in R 3.2 (R Core Team 2013). For all microsatellite analyses, loci that appeared to have null alleles were removed from each species, with Hr13 removed from *H.* spp. B and D, and Hr1 removed from *H.* spp. B and C. Genetic differentiation was quantified by pairwise F_{ST} (Weir and Cockerham 1984) and G''_{ST} (Hedrick 2005) using DIVERSITY. G''_{ST} is a standardized measure of genetic differentiation that accounts for levels of genetic variability, allowing the comparison of differentiation among loci and organisms with different effective population sizes (Hedrick 2005). For the mitochondrial data, haplotype diversity (h) and nucleotide diversity (π), average number of pairwise differences were calculated using DNASP 5.10 (Librado and Rozas 2009). Analyses of Molecular Variance (AMOVA) were utilized for hierarchical analysis of the geographical pattern of genetic

differentiation, with ARLEQUIN 3.5 and GENALEX 6.5 for mitochondrial and microsatellite data, respectively.

Discriminant analyses of principal components (DAPC; Jombart *et al.* 2010) were performed using the ADEGENET package in R (R Core Team 2013). The number of principal components was selected by evaluating the number conferring the lowest mean squared error and the highest mean successful assignment. Where these values did not concord, the mean value of the two measures was used. The K-means clustering procedure was used to identify the number of clusters in the data after transformation using principal component analysis in ADEGENET.

Spatial structure analyses were performed in GENALEX 6.5 to determine whether there are patterns of isolation by distance. The autocorrelation coefficient *r*, representing genotypic similarity, is compared at defined geographical distance classes. To further reveal the extent of genetic structure, *r* was calculated for increasing distances classes spanning the maximum distance between colonies. Because this method utilizes genotypic data, it sensitive to more finescale processes and shorter timescales than allele-frequency based analyses. Comparisons of spatial autocorrelation between males and females were conducted to examine evidence for sex-biased dispersal.

Acoustic analysis

Echolocation calls of stationary (hand-held) bats were recorded using a handheld ultrasound detector (Pettersson D1000X; Pettersson Elektronik AB, Uppsala, Sweden; sampling rate 400 kHz). Measurements of peak frequency of the constant frequency (CF) component of calls were extracted from acoustic files following methods outlined in Chapter 5. We performed Kruskal–Wallis tests on differences in peak call frequency of adult bats by species according to location (due to low sample sizes in B and C, colonies were pooled to location for all species). Post-hoc Mann-Whitney-Wilcoxon tests were then performed for pairwise comparisons.

RESULTS

Genetic diversity

Of 379 cyt *b* sequences, we identified 160 polymorphic sites (166 total mutations) corresponding to 57 unique haplotypes. *Hipposideros* sp. D was identified at all bat colonies sampled, while *H.* sp. C was observed at all but Likpe 1 and Forikrom. *Hipposideros* sp. B was found only in the Techiman region (Buoyem, Forikrom). Observed and expected heterozygosity and F_{IS} for microsatellites and summary statistics are found in table 2. Diversity statistics for mitochondrial data are found in table 3.

Table 2. Genetic diversity from microsatellite loci by population

Species	*n*	mean *n*	# alleles	AR	*Ho*	*He*	F_{IS}
***H.* sp. B**							
Buoyem 1	24	20.9	75	2.29	0.63	0.70	0.109
Buoyem 2	18	17.6	61	2.59	0.67	0.74	0.046
***H.* sp. C**							
Akpafu	19	17.3	82	4.91	0.65	0.72	0.098
Buoyem 1	15	13.9	69	4.63	0.65	0.74	0.124
Kwamang	7	6.6	63	4.59	0.70	0.74	0.053
Likpe 2	15	14.3	68	4.59	0.70	0.74	0.164
***H.* sp. D**							
Buoyem 2	23	21.6	72	4.46	0.68	0.70	0.025
Forikrom	24	23.3	71	4.43	0.67	0.68	0.007
Kwamang 1	39	37.7	94	5.23	0.73	0.77	0.057
Kwamang 2	40	38.3	83	4.87	0.71	0.75	0.061
Kwamang 3	7	6.6	60	4.49	0.8	0.74	-0.083
Likpe 1	23	22.1	65	4.17	0.72	0.70	-0.024
Likpe 2	18	17.6	63	4.33	0.72	0.72	-0.007

Calculated where $n \geq 5$. *p*-values for HWE by population were not significant after Bonferroni correction, with the exception of Akpafu ($p < 0.0001$). Number of loci = 10 (species B), 11 (species C and D). *n* = number of individuals included, mean *n* = mean number of individuals typed per locus.

Phylogeographic analyses

Bayesian phylogenetic analyses reveal a phylogeny that is consistent with four monophyletic species (A, B, C, and D; figure 1c). For the three species of interest, Ghanaian samples are highly divergent from other major intraspecific lineages found elsewhere (lineages C3, B3, and D3; Chapter 5). In Ghana, C and D were each separated into two well-supported clades, which were clearly geographically delineated. D was separated into eastern and western clades, while for C,

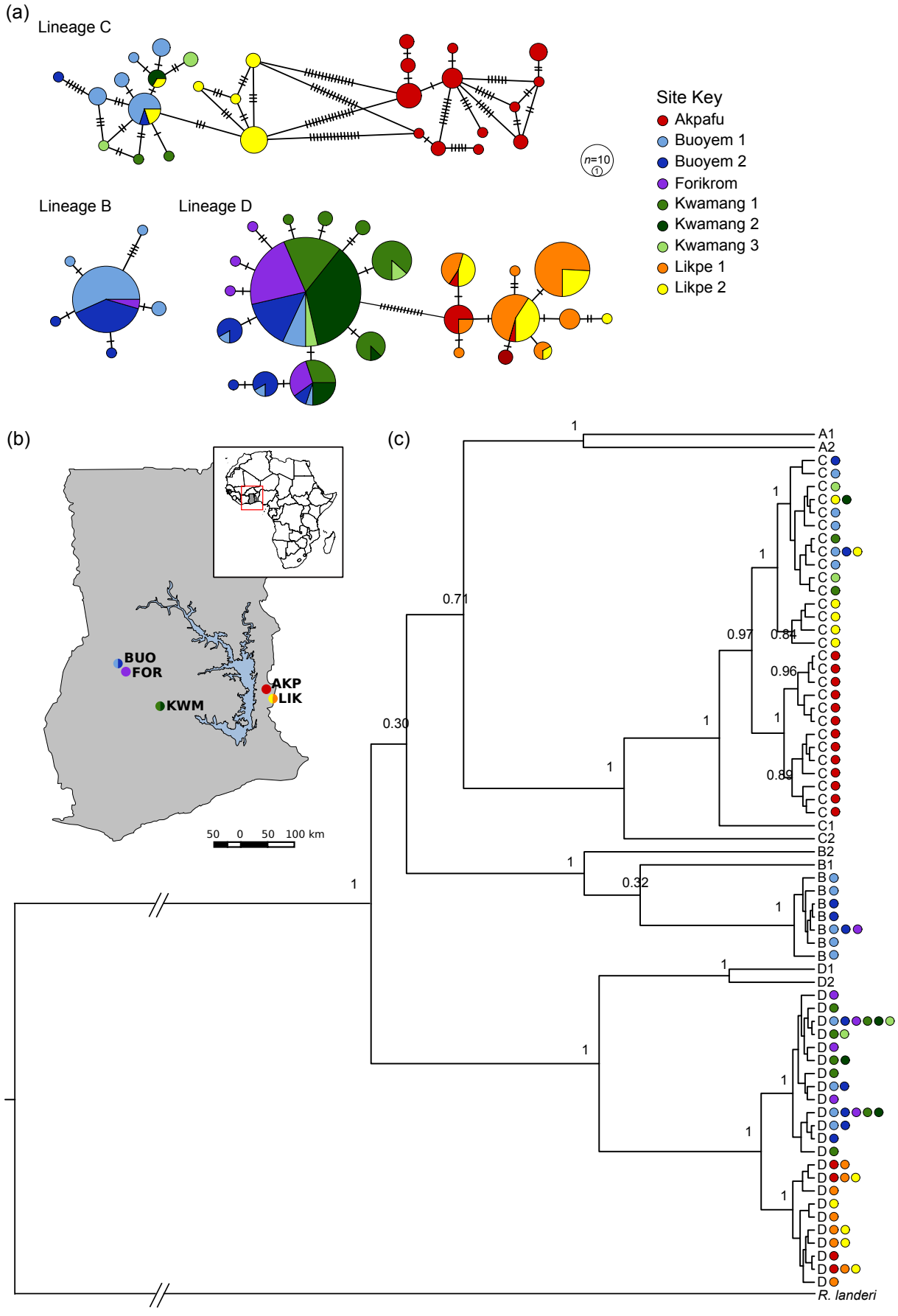
Akpafu formed its own distinct clade while Likpe in the east was clustered with the western sites. Within this 'western' clade was a highly supported clade found only at Likpe. For *H.* sp. D, there was a larger proportion of shared haplotypes among populations than for the other species. These findings are reflected in the minimum-spanning network (figure 1a).

Pairwise genetic divergences (K2P) were 1.2 – 3.7% between the C clade comprising samples from Akpafu and the clade representing all other sites. Within the latter subclade, K2P distances between the monophyletic Likpe haplotypes and the western haplotypes were 0.4 – 2%. Between the western and eastern D subclades, K2P distances were 1.2 – 2.1%. Taking the rate of sequence evolution at the tree tips as approximately 4% (Chapter 5) and the K2P distance ranges, this places most recent common ancestors for these intraspecific clades at roughly 0.3 - 0.52 Ma for the western and eastern D subclades, and 0.3 – 0.93 Ma for the two major C subclades identified at our study sites. Within *H.* sp. C, the Likpe group diverged from the western haplotypes approximately 0.1 – 0.5 Ma.

Demographic analyses

The star-like haplotype network for *H.* sp. B suggests a recent population expansion, while the network for *H.* sp. C supports a constant historical population size in *H.* sp. C (figure 1a). *Hipposideros* sp. D shows evidence of recent population expansion for the western sites (Buoyem, Kwamang, Forikrom), while the eastern clade shows a pattern consistent with stability.

Figure 1. (opposite) a) Minimum spanning networks for *H.* sp. B (*n*=50), C (*n*=66), and D (*n*=263). Circle size is proportional to haplotype frequency. b) Map of sampling locations, BUO = Buoyem, KWM = Kwamang, FOR = Forikrom, AKP = Akpafu-Todzi (hereafter 'Akpafu'), LIK = Likpe-Todome (hereafter 'Likpe'). c) Bayesian consensus phylogram with support for the major nodes. GenBank accession numbers were as follows: A1 (EU934452.1), A2 (FJ347977), B1 (EU934474.1), B2 (EU934462.1), C1 (FJ347994.1), C2 (FJ347996.1), D1 (HQ343258.1), D2 (HQ343248.1), *Rhinolophus landeri* (FJ457612.1).


(a)
Lineage C
Lineage B
Lineage D
n=10
Site Key
Akpafu
Buoyem 1
Buoyem 2
Forikrom
Kwamang 1
Kwamang 2
Kwamang 3
Likpe 1
Likpe 2
(b)
BUO
FOR
KWM
AKP
LIK
50 0 50 100 km
(c)
A1
A2
C1
C2
B2
B1
D1
D2
R. landeri

These results are supported by neutrality statistics and mismatch distributions. The significant, large negative F_S, F^* and D^* for *H.* sp. B are indicative of recent population expansion (table 3). This is supported by a large expansion coefficient and a unimodal mismatch distribution, as expected with a recent expansion. The very low estimate of τ (0.000) from the mismatch distribution for *H.* sp. B suggests that this population expansion occurred quite recently. *Hipposideros* sp. C and D showed multimodal and bimodal mismatch distributions, respectively, and non-significant neutrality statistics, indicative of constant historical population size, however, the western clade of *H.* sp. D alone shows significant, large negative F_S, a large expansion statistic and a unimodal mismatch distribution, and a τ of 0.000, indicating a very recent population expansion. Raggedness statistics were not significant. The relatively high expansion statistics suggest some level of population growth for all groups. For a very crude maximum estimate of the timing of the population expansion for *H.* sp. D, we used the calculation of $\tau = 2u$, where u = mutation rate in millions of years. We used the overall estimate of τ calculated for *H.* sp. D. The values of $\tau = 0.000$ estimated for the *H.* sp. D western clade and *H.* sp. B prohibit further calculation. Given a lineage-wise mutation rate of approximately 4% per million years (calculated based on mutation rate of *H.* sp. B in Bayesian phylogenetic tree, Chapter 5), and a generation time of two years (H. Baldwin unpublished data; Brosset 1968; Brosset 1969), this gives an estimate of approximate date of expansion of later than 14 000 years BP.

Significant heterozygote excesses for *H.* sp. B were observed using the Wilcoxon signed-rank test for the infinite allele ($p < 0.001$; $p < 0.001$) and two-phase ($p < 0.01$; $p < 0.001$) mutation models, and using the sign test under the infinite allele ($p = 0.01$; $p < 0.01$) for Buoyem 1 and Buoyem 2, respectively. Using the sign test under the two phase model, heterozygote excess was significant Buoyem 2. All tests were non-significant under the stepwise mutation model. There was no consistent or strong pattern of significant heterozygote excess for the western *H.* sp. D group. Allele frequency distribution did not differ from the L-shaped curve for either group.

Table 3. Diversity statistics, neutrality tests, and mismatch distribution analysis. Results are given for *H.* spp. B, C, and D, and separately for the western and eastern clades of *H.* sp. D.

	***H.* sp. B**	***H.* sp. C**	***H.* sp. D**	***H.* sp. D (west)**	***H.* sp. D (east)**
Nucleotide diversity (π)	0.0005	0.0151	0.0088	0.0021	0.0018
Avg pairwise nt diffs (*d*)	0.358	11.829	6.893	1.607	1.4288
Haplotype diversity (*h*)	0.227	0.951	0.772	0.589	0.780
n (# haplotypes)	50 (6)	66 (28)	263 (23)	187(16)	82(10)
Expansion coefficient (*S/d)*	22.35	4.73	3.34	18.05	7.70
Fu's (1997) F_S	-4.486**	-2.354	1.230	-5.645*	-2.550
Fu and Li's (1993) *F**	-3.770**	-0.507	-0.506	-1.225	-1.758
Fu and Li's (1993) *D**	-3.694**	-0.663	-1.058	-0.386	-1.765
Raggedness (*rg*)	0.384	0.010	0.051	0.0727	0.0497
Mismatch distribution	Unimodal	Multimodal	Bimodal	Unimodal	Unimodal
τ	0.000	4.514	0.429	0.000	1.429

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Population genetic analyses

Measures of genetic structure were calculated for all samples of five or more bats per site. Due to low sample sizes from Forikrom, population genetic analyses were limited to comparisons between the two closely situated Buoyem colonies. Kwamang sites were pooled for *H.* sp. C in order to increase sample size. Hierarchical analysis of molecular variance (AMOVA) and pairwise F_{ST} reveal contrasting levels of genetic subdivision among sampling localities (tables 4, 5, 6, 7; figure 2). Significant genetic structure was observed from microsatellite data for *H.* sp. C (Φ_{PT} =0.144, p = 0.001, global F_{ST} =0.094, p = 0.001) and *H.* sp. D (Φ_{PT} = 0.04, p = 0.001, global F_{ST} = 0.049, p =0.001). Unsurprisingly, *H.* sp. B showed no differentiation between the two closely situated colonies with adequate sampling size (Φ_{PT} = 0.039, p = 0.158; global F_{ST} = 0.024, p = 0.136). Global Φ_{ST}s, based on mitochondrial data, were 0.154 for *H.* sp. C, 0.283 for D, and -0.015 for B based on conventional haplotype frequencies, and 0.773 for C, 0.885 for D, and 0.000 for B based on pairwise difference (p = 0.00000 for all measures for C and D, not significant for *H.* sp. B) (Excoffier *et al.* 1992; Weir and Cockerham 1996; Weir and Cockerham 1984).

Table 4. Conventional Φ_{ST} for *H.* sp. C based on cytochrome b data (n = 66).

	AT1	BUO1	KW
AT1	—		
BUO1	0.15***	—	
KW	0.09**	0.17**	—
LT2	0.16***	0.19***	0.17***

*p < 0.05, **p < 0.01, ***p < 0.001

Table 5. F_{ST} (below diagonal) G''_{ST} (above diagonal) and for *H.* sp. C based on microsatellite data (n = 56).

	AT1	BUO1	KW	LT2
AT1	—	0.4885*	0.5449*	0.5596*
BUO1	0.122***	—	0.0933	0.339*
KW	0.127***	0.014	—	0.2458*
LT2	0.135***	0.075***	0.045***	—

F_{ST}: *p < 0.05, **p < 0.01, ***p < 0.001 (Fisher's exact method)

G''_{ST}: * statistically significant (CIs do not overlap 0)

Table 6. Φ_{ST} for *H.* sp. D based on cytochrome b data (n = 263).

	AT1	BUO1	BUO2	FO	KW1	KW2	KW3	LT1
AT1	—							
BUO1	0.43***	—						
BUO2	0.36***	-0.03	—					
FO	0.50***	-0.03	0.06*	—				
KW1	0.28***	0.09*	0.09***	0.13***	—			
KW2	0.66***	0.02	0.15***	0.02	0.23***	—		
KW3	0.40***	0.01	0.04	0.07	-0.01	0.19*	—	
LT1	0.26***	0.36***	0.31***	0.41***	0.25***	0.52***	0.33***	—
LT2	0.26***	0.37***	0.32***	0.43***	0.25***	0.57***	0.34***	0.01

*p < 0.05, **p < 0.01, ***p < 0.001

Table 7. F_{ST} (below diagonal) G''_{ST} (above diagonal) and for *H.* sp. D based on microsatellite data (n = 174).

	BUO2	FO	KW1	KW2	KW3	LT1	LT2
BUO2	—	0.0065	0.171*	0.144*	0.131*	0.239*	0.215*
FO	0.001	—	0.194*	0.151*	0.154*	0.253*	0.268*
KW1	0.047***	0.055***	—	0.048	-0.043	0.181*	0.143*
KW2	0.041***	0.045***	0.011***	—	-0.032	0.152*	0.126*
KW3	0.037***	0.046***	-0.013	-0.010	—	0.058	0.025
LT1	0.076***	0.084***	0.049***	0.043***	0.018**	—	0.008
LT2	0.066***	0.085***	0.036***	0.034***	0.007	0.002	—

F_{ST}: *p < 0.05, **p < 0.01, ***p < 0.001 (Fisher's exact method)

G''_{ST}: * statistically significant (CIs do not overlap 0)

There was generally little genetic differentiation between colonies that were very close (<5 km, exception: mtDNA between KW1 and KW2 for *H.* sp. D) (tables 4, 5, 6, 7; figures 3, 4). However, there was significant genetic structuring even at very small spatial scales for *H.* sp. C and D (e.g. Likpe and Akpafu, ~17km).

DAPC scatterplots are shown in figure 2. Three distinct clusters can be identified for *H.* sp. C, with no overlap of 95% confidence ellipses between Akpafu, Buoyem 1, and a third group comprising Kwamang and Likpe 2, while K-means clustering identified two clusters, grouping Akpafu separately from the other three colonies (figure 2b). Three K-means clusters are identified for *H.* sp. D, which roughly correspond to Buoyem and Forikrom, Kwamang, and Likpe (figure 2c). For *H.* sp. B, DAPC clusters largely overlap and a single K-means cluster was identified (figure 2a).

Spatial structure autocorrelation showed significantly higher genotypic similarity within colonies for *H.* sp. C and D, declining with geographic distance (figure 3). *H.* sp. C exhibited oscillation of high and low levels of autocorrelation, reflecting the higher genetic structure between the Akpafu site and all others, and slightly but significantly higher than expected relatedness between Kwamang and Buoyem (figure 3a). This pattern is consistent with localised movement and the effects of historical genetic drift. For *H.* sp. D, we observe significantly higher relatedness within colonies, with a decline in relatedness between colonies at increasing distances apart, and significantly lower relatedness at the highest geographic distance, a clear pattern of isolation by distance (figure 3b). Similarly, analysis of the autocorrelation coefficient *r* for increasing distance size classes showed significantly higher genotypic similarity at shorter distance class bins, with a decline in relatedness with distance for *H.* sp. C and D (figure 4). A sharp drop in *r* is observed for *H.* sp. C at distances greater than 4km, reflecting high relatedness between the Kwamang sites, while for *H.* sp. D the decline is gradual. Due to being limited to two sites with adequate sampling number, spatial structure analyses were not performed for *H.* sp. B. D (figure

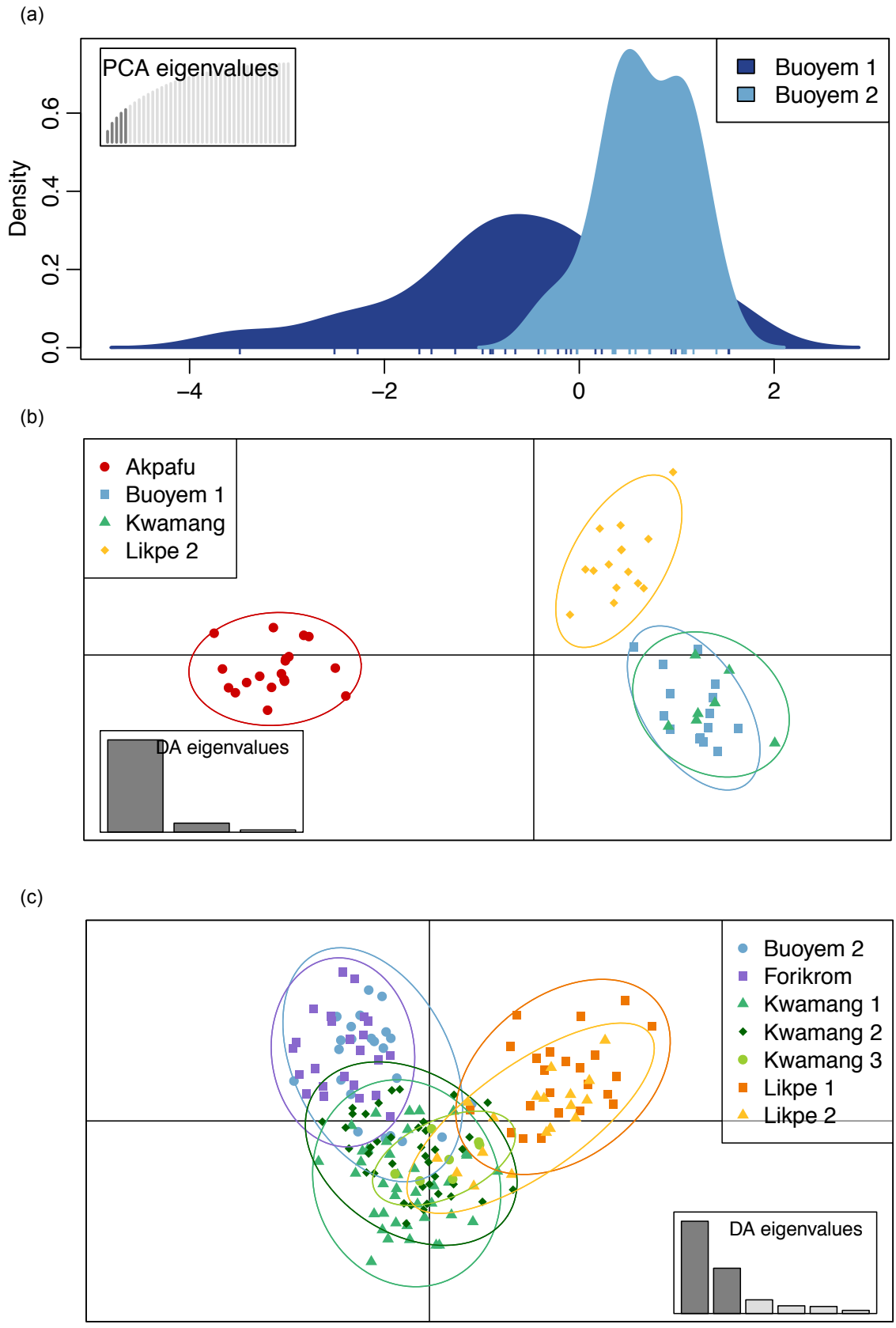


Figure 2. DAPC plots for *H.* spp. a) B, b) C, and c) D. Number of axes retained in discriminant analysis = 1, 3, and 6, respectively.

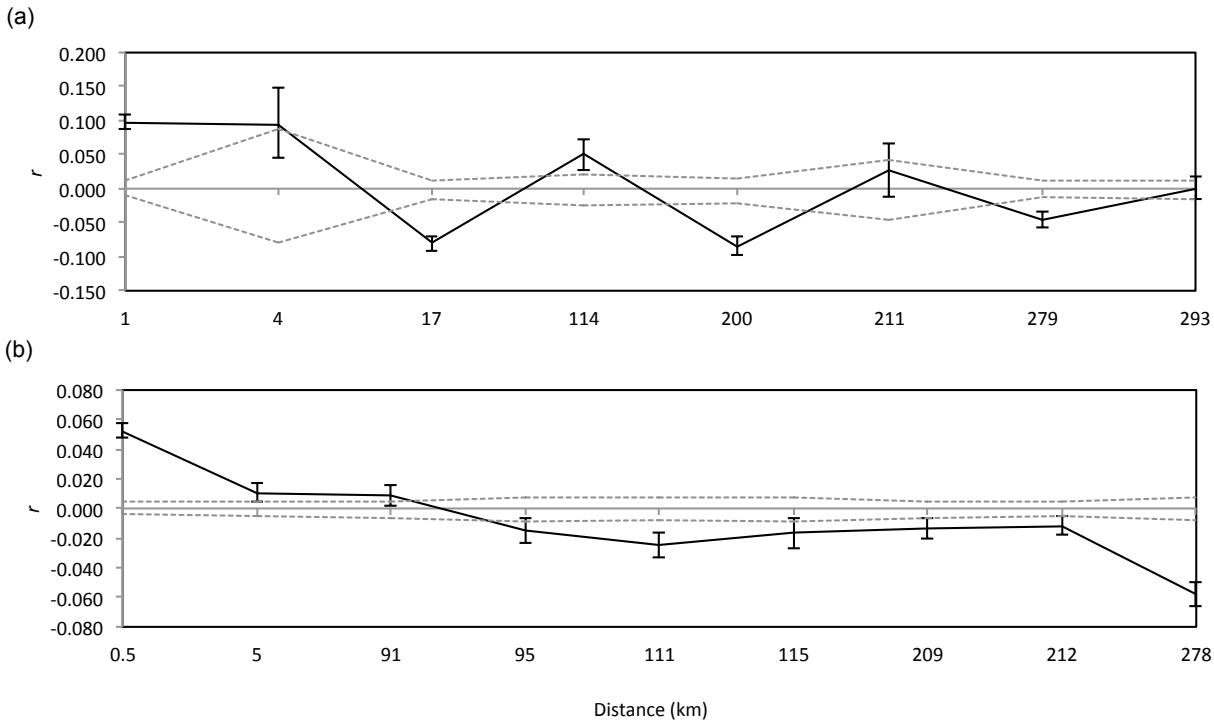


Figure 3. Spatial autocorrelogram for a) *H.* sp. C, b) *H.* sp. D.

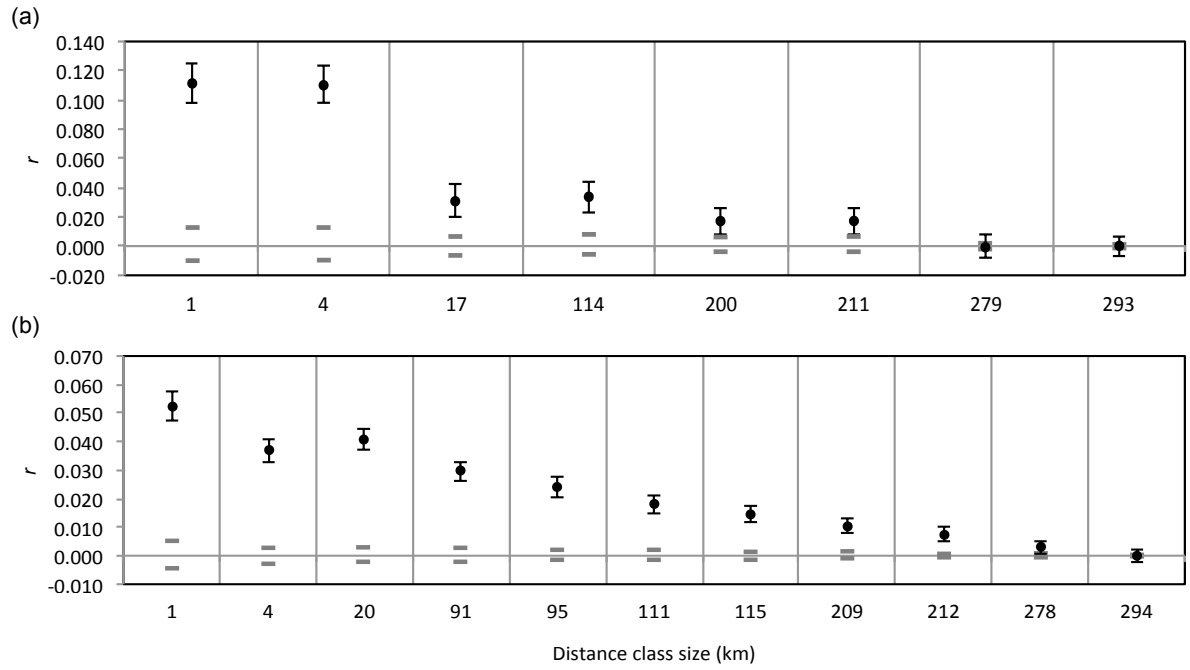


Figure 4. Spatial autocorrelation with increasing distance class bins for a) *H.* sp. C, and b) *H.* sp. D. Circles = *r* (relatedness), error bars bound the 95% CI about r as determined by bootstrap resampling (999 permutations, 1000 bootstraps), grey bars = CI for null hypothesis (*r* = 0)

4). There were no differences in patterns of autocorrelation between males and females (figure S1).

Acoustic analysis

Hipposideros sp. C showed significant differences in echolocation call peak frequency among locations, with bats captured at Akpafu calling at frequencies of approximately 10 – 11.5 kHz higher than those sampled at the other colonies and locations (table 8, Kruskal-Wallis test $p < 0.0001$, Mann-Whitney-Wilcoxon test $p < 0.01$ for all pairwise comparisons including Akpafu, not significant for all other comparisons (supplementary material). There were no significant differences in mean echolocation call frequency between locations for *H.* sp. D, and sample sizes were too low to statistically test for *H.* sp. B.

Table 8. Differences in call peak frequency by location.

	***H.* sp. B**			***H.* sp. C**			***H.* sp. D**		
Location	*n*	Mean PF (kHz)	95% CI	*n*	Mean PF (kHz)	95% CI	*n*	Mean PF (kHz)	95% CI
Akpafu	0	—	—	12	147.5	(141.7–153.2)	1	132.0	—
Buoyem	16	125.2	(122.6–127.9)	12	136.2	(129.6–142.9)	15	132.3	(126.9–137.8)
Kwamang	0	—	—	4	135.9	(129.5–142.3)	41	133.7	(129.5–137.1)
Likpe	0	—	—	6	136.3	(133.0-139.6)	27	132.9	(128.0–139.3)
Forikrom	1	127.1	—	0	—	—	26	133.3	(128.9–136.8)

DISCUSSION

Overall, our results show that despite co-distribution and similar ecological and natural history traits, the recently diverged bat species *Hipposideros* spp. B, C, and D in West Africa show surprisingly different genetic patterns. There are important differences in phylogeography and gene flow, patterns of isolation by distance, and interesting contrasts in demographic history. These findings suggest contrasting responses to the same historical and contemporary processes. The results presented here provide some insights into biogeographic processes of the Upper Guinean tropical forests region, and the variation in species responses to extrinsic forces.

Phylogeny and phylogeography

Hipposideros sp. C and D exhibited strong similarity in phylogenetic patterns, each comprising two reciprocally monophyletic subclades with similar levels of divergence. In contrast, *H.* sp. B was found exclusively in western sites, and showed a single clade with comparatively few haplotypes. Phylogeographical patterns contrast with the phylogenetic similarity in *H.* sp. D and C. The major clades of *H.* sp. D are allopatric and suggest a relatively ancient split separating the eastern and western sites, while *H.* sp. C shows a different pattern, with an ancient split between a single site in the east, and the remaining sites, with some shared haplotypes among eastern and western sites. The patterns observed within this latter group suggest that the population at Likpe was founded by dispersal from the west. Different patterns of colonisation may contribute to these observed differences.

Although the patterns are different, it is possible to identify varying levels of population genetic subdivision into eastern and western groups. This pattern can be interpreted as evidence of past vicariant isolation, consistent with historical allopatry, such as isolation of eastern and western populations due to forest contraction during the Pleistocene followed by range expansion and secondary contact within these species (Crandall *et al.* 2008). The deeply split clades and heterogeneity of divergence shown in phylogenetic analyses and haplotype networks in these species may reflect the cyclical nature of Pleistocene climatic change and the resulting forest contraction and expansion which may have differentially affected lineage divergence among co-distributed species (Dynesius and Jansson 2000). Pleistocene conditions resulting in intermittent isolation, and secondary contact and introgression are also suggested to have shaped similar patterns of highly divergent, geographically structured haplotype groups in previous molecular studies of African vertebrates (e.g. Barlow *et al.* 2013; Bell *et al.* 2011; Blackburn 2008; Brouat *et al.* 2009; Hassanin *et al.* 2015; Leaché and Fujita 2010; Nicolas *et al.* 2008), and hipposiderid and related bats (Chen *et al.* 2006; Lin *et al.* 2014; Rossiter *et al.* 2007).

Population demographic history

Diversity and neutrality statistics and mismatch distribution analyses revealed contrasting population demographic histories. *Hipposideros* sp. C appears to have a long evolutionary history of a large, stable population, whereas *H.* sp. B appears to have experienced a recent population expansion, resulting either from a population bottleneck or founder event. *Hipposideros* sp. D shows contrasting population histories for the eastern and western clades, with a very recent expansion in the west and a historically stable eastern population. Given the lack of shared alleles with the eastern haplotypes, colonisation from further west of the range seems more plausible than a bottleneck. Timing of these population expansions could not be calculated because the value of tau for both rounded to zero, indicating very recent expansion events. Therefore, while these patterns could be related to habitat loss from forest contraction during the quaternary glaciations resulting in the local extinctions and recolonisations (Crandall *et al.* 2008), the very low estimate of tau suggests that the expansions occurred well after the last glacial maximum.

Genetic diversity and connectivity

Genetic diversity is highest in *H.* sp. C, consistent with a large, constant historical population size indicated by the mismatch distribution, diversity statistics and haplotype network. This is in contrast to the single, far less genetically diverse *H.* sp. B, which shows relatively depauperate mitochondrial genetic diversity, even after accounting for the finer geographic scale of sampling, and lower allelic richness. We find a surprisingly high level of genetic partitioning between colonies and isolation by distance at relatively fine scales (<350km). Levels of gene flow were higher for *H.* sp. C than D, with no differentiation for *H.* sp. B, which is unsurprising given the close proximity of the two colonies that could be included in the analysis. A clear pattern of isolation by distance was observed in *H.* sp. D, in contrast to *H.* sp. C, which showed structure more consistent with localised movement patterns and a sharp discontinuity by distance, with evidence of drift retained from colonisation of western-derived lineages at Likpe. The difference

in echolocation call frequency between *H.* sp. C at Akpafu and the other sites may be interpreted as the continued effects of drift resulting from past vicariant isolation after the apparent recent colonisation of the Likpe from the western clade. Further, this acoustic divergence may affect communication and mate choice between these clades (Chapter 5; Jones and Barlow 2004; Kingston *et al.* 2001), further explaining the genetic structuring between these two relatively close colonies.

Mitochondrial and microsatellite data showed consistent genetic differentiation among sites, although levels of structure were consistently higher for mitochondrial DNA. Different patterns of spatial autocorrelation between males and females were not observed in the nuclear data, and therefore this does not likely reflect sex-biased dispersal, and is probably a reflection on lower effective population sized for mitochondrial than nuclear markers.

High levels of genetic structure and/or isolation by distance have been observed in other hipposiderid bats (Echenique-Díaz *et al.* 2009; Lin *et al.* 2014; Xu *et al.* 2010), closely related rhinolophid bats (Chen *et al.* 2006; Rossiter *et al.* 2000) and other bat species (for review see Moussy *et al.* 2012). Hipposiderids are cluttered space foragers, with low aspect ratio and wing loading and rounded wing tips, associated with limited dispersal ability (Norberg and Rayner 1987) and relatively small home ranges (Nkrumah in prep). The naturally uneven distribution of cave roosts, combined with a highly modified agricultural landscape that may limit the availability of tree roosts, may also contribute to these patterns (Meyer *et al.* 2009).

The eastern sites of Upper Guinean forests and the western sites in the moist, elevated forest island of the Togo Hills are separated by the dry savannah of the Dahomey Gap (Leaché and Fujita 2010; Murienne *et al.* 2013), which dates to the Holocene, approximately 3400-4500 years BP (Salzmann and Hoelzmann 2005), and may be a significant barrier to dispersal. The Dahomey Gap has been implicated in promoting intraspecific genetic divergence in other vertebrates (e.g. Blackburn 2008; Leaché and Fujita 2010; Nicolas *et al.* 2008). The Volta River is another potential dispersal barrier, which has been suggested for other small vertebrates (Brouat

et al. 2009; Jacquet *et al.* 2014; Nicolas *et al.* 2008). Several studies have shown that some bat species avoid flying over open water (see review in Moussy *et al.* 2012), including water bodies as narrow as <180m (Albrecht *et al.* 2007; Meyer *et al.* 2009). Hipposiderid bats prefer to fly in cluttered forest understory, and may be unlikely to cross rivers and lakes (Echenique-Díaz *et al.* 2009).

Implications

The recent recognition of new species within this group of bats (Chapter 5; Vallo *et al.* 2008) means that abundances have been historically overestimated, and the true range distributions of these species are not well understood. The true conservation statuses of these bats remain, therefore, unknown. This study revealed low levels of genetic diversity in *H.* sp. B, and potentially more sensitive habitat specificity given its absence at sites where *H.* sp. C and D were present within its geographic range. Further work is needed to assess its conservation status with broader geographic sampling. Further, the genetic structuring observed for *H.* sp. C and isolation by distance in *H.* sp. D on the relatively fine geographic scale of this study may imply high susceptibility to habitat fragmentation. As West Africa is a hotspot for anthropogenic habitat destruction (Malhi *et al.* 2013), and bat species in this region are exposed to many potential threatening processes such as loss of foraging habitat and roosting sites due to logging, forest burning and habitat encroachment (Mickleburgh *et al.* 2009), this warrants further investigation. Further, these insights into dispersal provide information useful for modelling predictions about landscape disease transmission, which may inform management strategies in the event of future outbreaks of CoVs from hipposiderid bats in Africa.

The geographically discordant patterns of genetic differentiation found in this study suggest that these taxa have responded differently to the same historical biogeographic processes (Avise 2009; McGovern *et al.* 2010). Different responses to extrinsic forces, have also been observed via discordant patterns of genetic structure in ecologically similar, congeneric

butterflyfish (DiBattista *et al.* 2012) and marine gastropods (Crandall *et al.* 2008). In spite of the broad similarities in the ecologies and natural histories of these bats, subtle differences, such as in macro- or microhabitat preference (Paz *et al.* 2015), social behaviour (Chen *et al.* 2010; Lange *et al.* 2010; Whiteley *et al.* 2004), or extent of physiological tolerances (Worthington-Wilmer *et al.* 1999; Worthington-Wilmer *et al.* 1994), may contribute to these contrasting responses. Our findings suggest that even with very high levels of ecological similarity in closely related, sympatric taxa, responses to extrinsic processes may be substantially different.

CONCLUSIONS

The contrast in contemporary gene flow, dispersal, and demographic histories among bats with similar ecological and natural history traits, and presumably the same long-term biogeographical history, was unexpected. Our findings show that species-specific information is crucial to our understanding of patterns of intraspecific genetic diversity in these taxa. Using one species to predict patterns of diversity or demography would be inappropriate for informing management decisions about all three. These findings highlight the need to be wary of inferring intraspecific genetic patterns from data collected from closely related or similar species.

ACKNOWLEDGEMENTS

We thank everybody who was involved in field work; Isaac Mawusi Adanyeguh, Lucinda Kirkpatrick, Anna Vogeler, Mac Elikem Nutsuakor, David Ofori Agyei, Sarah Koschnicke, Julia Morrison, Emmanual Asare, Eunice Okyere, Kenneth Quansah, Emmanuel Essoun, Julian Schmid, Florian Gloza-Rausch, Maximilian Vollstaedt, Ebenezer Gyimah, Paul Marfo, Kennedy Darkwa, Justice Konadu, Paul, Kwame Takye, Winfred, Samuel, Francis and Thomas Kruppa. We are indebted to the communities, community leaders and administrators of the sampling sites, the helpful staff of the Ghana Wildlife Division, and support staff at the Kumasi Centre for

Collaborative Research in Tropical Medicine. Thank you to Stephen Hoggard and Toni Mizerek, for helpful comments on the manuscript.

REFERENCES

Albrecht L., Meyer C. F. J. & Kalko E. K. V. (2007) Differential mobility in two small phyllostomid bats, *Artibeus watsoni* and *Micronycteris microtis*, in a fragmented neotropical landscape. *Acta Theriologica* 52(2), 141-149.

Altizer S., Harvell D. & Friedle E. (2003) Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology & Evolution* 18(11), 589-596.

Annan A., Baldwin H., Corman V., Klose S., Owusu M., Nkrumah E., Badu E., Anti P., Agbenyega O., Meyer B., Oppong S., Sarkodie Y., Kalko E., Lina P., Godlevska E., Reusken C. B. E. M., Seebens A., Gloza-Rauch F., Vallo P., Tschapka M., Drosten C. & Drexler J. F. (2013) Human betacoronavirus 2c EMC/2012–related viruses in bats, Ghana and Europe. *Emerging Infectious Diseases* 19(3), 456-459.

Archie E., Luikart G. & Ezenwa V. (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution* 24(1), 21-30.

Avise J. C. (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography* 36(1), 3-15.

Baldwin H., Vallo P., Gardner M., Drosten C., Tschapka M. & Stow A. (2014) Isolation and characterization of 11 novel microsatellite loci in a West African leaf-nosed bat, *Hipposideros* aff. *ruber*. *BMC Research Notes* 7(1), 607-610.

Barlow A., Baker K., Hendry C. R., Peppin L., Phelps T., Tolley K. A., Wuester C. E. & Wuester W. (2013) Phylogeography of the widespread African puff adder (*Bitis arietans*) reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology* 22(4), 1134-1157.

Bell G. & Fenton M. (1983) Echolocation and foraging behavior of *Hipposideros ruber* (Chiroptera). *The Journal of the Acoustical Society of America* 74, S32.

Bell G. P. & Fenton M. B. (1984) The use of Doppler-shifted echoes as a flutter detection and clutter rejection system: the echolocation and feeding behavior of *Hipposideros ruber* (Chiroptera: Hipposideridae). *Behavioral Ecology and Sociobiology* 15(2), 109-114.

Bell R. C., MacKenzie J. B., Hickerson M. J., Chavarría K. L., Cunningham M., Williams S. & Moritz C. (2011) Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. *Proceedings of the Royal Society of London B: Biological Sciences* 279(1730), 991-999.

Biek R. & Real L. A. (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology* 19(17), 3515-3531.

Blackburn D. C. (2008) Biogeography and evolution of body size and life history of African frogs: phylogeny of squeakers (*Arthroleptis*) and long-fingered frogs (*Cardioglossa*) estimated from mitochondrial data. *Molecular Phylogenetics and Evolution* 49(3), 806-826.

Brosset A. (1968) *La Permutation du Cycle sexuel saisonnier chez le chiroptère Hipposideroscaffer, au voisinage de l'Équateur*.

Brosset A. (1969) Recherches sur la biologie des Chiroptères troglophiles dans le nord-est du Gabon. *Biologie du Gabon* 5, 93-116.

Brouat C., Tatard C., Bâ K., Cosson J. F., Dobigny G., Fichet-Calvet E., Granjon L., Lecompte E., Loiseau A. & Mouline K. (2009) Phylogeography of the Guinea multimammate mouse (*Mastomys erythroleucus*): a case study for Sahelian species in West Africa. *Journal of Biogeography* 36(12), 2237-2250.

Calisher C. H., Childs J. E., Field H. E., Holmes K. V. & Schountz T. (2006) Bats: important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews* 19(3), 531-545.

Chen J., Rossiter S. J., Flanders J. R., Sun Y., Hua P., Miller-Butterworth C., Liu X., Rajan K. E. & Zhang S. (2010) Contrasting genetic structure in two co-distributed species of Old World fruit bat. *PLOS One* 5(11), e13903-e13903.

Chen S.-F., Rossiter S. J., Faulkes C. G. & Jones G. (2006) Population genetic structure and demographic history of the endemic Formosan lesser horseshoe bat (*Rhinolophus monoceros*). *Molecular Ecology* 15(6), 1643-1656.

Chittick B., Stoskopf M., Heil N., Levine J. & Law M. (2001) Evaluation of sandbar shiner as a surrogate for assessing health risks to the endangered Cape Fear shiner. *Journal of Aquatic Animal Health* 13(2), 86-95.

Cornuet J. M. & Luikart G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144(4), 2001-2014.

Crandall E. D., Frey M. A., Grosberg R. K. & Barber P. H. (2008) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology* 17(2), 611-626.

Darriba D., Taboada G. L., Doallo R. & Posada D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8), 772-772.

Daszak P., Cunningham A. & Hyatt A. (2000) Emerging infectious diseases of wildlife- threats to biodiversity and human health. *Science* 287(5452), 443-449.

Dawson M. N. (2012) Parallel phylogeographic structure in ecologically similar sympatric sister taxa. *Molecular Ecology* 21(4), 987-1004.

Dawson M. N. (2014) Natural experiments and meta-analyses in comparative phylogeography. *Journal of Biogeography* 41(1), 52-65.

DiBattista J. D., Rocha L. A., Craig M. T., Feldheim K. A. & Bowen B. W. (2012) Phylogeography of two closely related Indo-Pacific butterflyfishes reveals divergent evolutionary histories and discordant results from mtDNA and microsatellites. *Journal of Heredity* 103(5), 617-629.

Drew J. A. & Barber P. H. (2012) Comparative phylogeography in Fijian coral reef fishes: a multi-taxa approach towards marine reserve design. *PLOS One* 7(10), e47710.

Drummond A. & Rambaut A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7(1), 214-221.

Dunning D. & Krüger M. (1996) Predation upon moths by free-foraging *Hipposideros caffer*. *Journal of Mammalogy* 77(3), 708-715.

Dynesius M. & Jansson R. (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences* 97(16), 9115-9120.

Echenique-Díaz L. M., Yokoyama J., Takahashi O. & Kawata M. (2009) Genetic structure of island populations of the endangered bat *Hipposideros turpis turpis*: implications for conservation. *Population Ecology* 51(1), 153-160.

Excoffier L., Smouse P. E. & Quattro J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131(2), 479-491.

Fu Y.-X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147(2), 915-925.

Fu Y.-X. & Li W.-H. (1993) Statistical tests of neutrality of mutations. *Genetics* 133(3), 693-709.

Guivier E., Galan M., Chaval Y., Xuéreb A., Ribas Salvador A., Poulle M. L., Voutilainen L., Henttonen H., Charbonnel N. & Cosson J. F. (2011) Landscape genetics highlights the role of bank vole metapopulation dynamics in the epidemiology of Puumala hantavirus. *Molecular Ecology* 20(17), 3569-3583.

Gutiérrez-García T. A. & Vázquez-Domínguez E. (2011) Comparative phylogeography: designing studies while surviving the process. *Bioscience* 61(11), 857-868.

Hassanin A., Khouider S., Gembu G.-C., Goodman S. M., Kadjo B., Nesi N., Pourrut X., Nakouné E. & Bonillo C. (2015) The comparative phylogeography of fruit bats of the tribe Scotonycterini (Chiroptera, Pteropodidae) reveals cryptic species diversity related to African Pleistocene forest refugia. *Comptes Rendus Biologies* 338(3), 197-211.

Hedrick P. W. (2005) A standardized genetic differentiation measure. *Evolution* 59(8), 1633-1638.

Jacquet F., Nicolas V., Colyn M., Kadjo B., Hutterer R., Decher J., Akpatou B., Cruaud C. & Denys C. (2014) Forest refugia and riverine barriers promote diversification in the West African pygmy shrew (*Crocidura obscurior* complex, Soricomorpha). *Zoologica Scripta* 43(2), 131-148.

Jombart T., Devillard S. & Balloux F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11(1), 94.

Jones G. & Barlow K. (2004) Cryptic species of echolocating bats. In: *Echolocation in bats and dolphins* (eds J. A. Thomas, C. F. Moss and M. Vater) pp. 345-349. University of Chicago Press, London.

Kingston T., Lara M., Jones G., Akbar Z., Kunz T. & Schneider C. (2001) Acoustic divergence in two cryptic *Hipposideros* species: a role for social selection? *Proceedings of the Royal Society Biological Sciences Series B* 268(1474), 1381-1386.

Lanfear R., Calcott B., Ho S. Y. & Guindon S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29(6), 1695-1701.

Lange R., Durka W., Holzhauer S. I., Wolters V. & Diekoetter T. (2010) Differential threshold effects of habitat fragmentation on gene flow in two widespread species of bush crickets. *Molecular Ecology* 19(22), 4936-4948.

Leaché A. D. & Fujita M. K. (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B: Biological Sciences* 277(1697), 3071-3077.

Librado P. & Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11), 1451-1452.

Lin A. Q., Csorba G., Li L. F., Jiang T. L., Lu G. J., Thong V. D., Soisook P., Sun K. P. & Feng J. (2014) Phylogeography of *Hipposideros armiger* (Chiroptera: Hipposideridae) in the oriental region: the contribution of multiple pleistocene glacial refugia and intrinsic factors to contemporary population genetic structure. *Journal of Biogeography* 41(2), 317-327.

Mace G. M. & Purvis A. (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Molecular Ecology* 17(1), 9-19.

Malhi Y., Adu-Bredu S., Asare R. A., Lewis S. L. & Mayaux P. (2013) African rainforests: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1625), 20120312.

McGovern T. M., Keever C. C., Saski C. A., Hart M. W. & Marko P. B. (2010) Divergence genetics analysis reveals historical population genetic processes leading to contrasting phylogeographic patterns in co-distributed species. *Molecular Ecology* 19(22), 5043-5060.

Meyer C., Kalko E. & Kerth G. (2009) Small scale fragmentation effects on local genetic Diversity in two phyllostomid bats with different dispersal abilities in Panama. *Biotropica* 41(1), 95-102.

Mickleburgh S., Waylen K. & Racey P. (2009) Bats as bushmeat: a global review. *Oryx* 43(2), 217-234.

Moritz C. (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology* 51(2), 238-254.

Moritz C. & Faith D. P. (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* 7(4), 419-429.

Moussy C., Hosken D., Mathews F., Smith G., Aegerter J. & Bearhop S. (2012) Migration and dispersal patterns of bats and their influence on genetic structure. *Mammal Review* 43, 183-195.

Murienne J., Benavides L. R., Prendini L., Hormiga G. & Giribet G. (2013) Forest refugia in Western and Central Africa as 'museums' of Mesozoic biodiversity. *Biology Letters* 9(1), 20120932.

Nicolas V., Mboumba J. F., Verheyen E., Denys C., Lecompte E., Olayemi A., Missoup A. D., Katuala P. & Colyn M. (2008) Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. *Journal of Biogeography* 35(11), 2074-2089.

Nkrumah E. E. (2014) Home range size of Noack's round-leaf bat (*Hipposideros* aff. *ruber*) in Ghana. PhD thesis. Kwame Nkrumah University of Science and Technology, Kumasi.

Norberg U. M. & Rayner J. (1987) Ecological morphology and flight in bats (Mammalia; Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 335-427.

Paz A., Ibáñez R., Lips K. R. & Crawford A. J. (2015) Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. *Molecular Ecology* 24, 3723-3737.

Peck D. R. & Congdon B. C. (2004) Reconciling historical processes and population structure in the sooty tern *Sterna fuscata*. *Journal of Avian Biology* 35(4), 327-335.

Pfefferle S., Oppong S., Drexler J., Gloza-Rausch F., Ipsen A., Seebens A., Müller M., Annan A., Vallo P., Adu-Sarkodie Y., Kruppa T. & Drosten C. (2009) Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerging Infectious Diseases* 15(9), 1377-1384.

Ponce-Reyes R., Clegg S. M., Carvalho S. B., McDonald-Madden E. & Possingham H. P. (2014) Geographical surrogates of genetic variation for selecting island populations for conservation. *Diversity and Distributions* 20(6), 640-651.

R Core Team. (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Real L. A. & Biek R. (2007) Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *Journal of the Royal Society Interface* 4(16), 935-948.

Rossiter S. J., Benda P., Dietz C., Zhang S. & Jones G. (2007) Rangewide phylogeography in the greater horseshoe bat inferred from microsatellites: implications for population history, taxonomy and conservation. *Molecular Ecology* 16(22), 4699-4714.

Rossiter S. J., Jones G., Ransome R. D. & Barratt E. M. (2000) Parentage, reproductive success and breeding behaviour in the greater horseshoe bat (*Rhinolophus ferrumequinum*). *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267(1443), 545.

Rouget M., Cowling R. M., Pressey R. L. & Richardson D. M. (2003) Identifying spatial components of ecological and evolutionary processes for regional conservation planning in the Cape Floristic Region, South Africa. *Diversity and Distributions* 9(3), 191-210.

Russell A., Medellin R. & McCracken G. (2005) Genetic variation and migration in the Mexican free tailed bat (*Tadarida brasiliensis mexicana*). *Molecular Ecology* 14(7), 2207-2222.

Salzmann U. & Hoelzmann P. (2005) The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *The Holocene* 15(2), 190-199.

Sappington L. C., Mayer F. L., Dwyer F. J., Buckler D. R., Jones J. R. & Ellersieck M. R. (2001) Contaminant sensitivity of threatened and endangered fishes compared to standard surrogate species. *Environmental Toxicology and Chemistry* 20(12), 2869-2876.

Schtickzelle N., WallisDeVries M. F. & Baguette M. (2005) Using surrogate data in population viability analysis: the case of the critically endangered cranberry fritillary butterfly. *Oikos* 109(1), 89-100.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30(12), 2725-2729.

Vallo P., Guillén-Servent A., Benda P., Pires D. & Koubek P. (2008) Variation of mitochondrial DNA in the *Hipposideros caffer* complex (Chiroptera: Hipposideridae) and its taxonomic implications. *Acta Chiropterologica* 10(2), 193-206.

Warman L. D., Forsyth D. M., Sinclair A., Freemark K., Moore H. D., Barrett T. W., Pressey R. L. & White D. (2004) Species distributions, surrogacy, and important conservation regions in Canada. *Ecology Letters* 7(5), 374-379.

Weir B. S. & Cockerham C. (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates, Sunderland, Massachusetts.

Weir B. S. & Cockerham C. C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.

Weir S. M., Suski J. G. & Salice C. J. (2010) Ecological risk of anthropogenic pollutants to reptiles: evaluating assumptions of sensitivity and exposure. *Environmental Pollution* 158(12), 3596-3606.

Whiteley A. R., Spruell P. & Allendorf F. W. (2004) Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. *Molecular Ecology* 13(12), 3675-3688.

Worthington-Wilmer J., Hall L., Barratt E. & Moritz C. (1999) Genetic structure and male-mediated gene flow in the ghost bat (*Macroderma gigas*). *Evolution*, 1582-1591.

Worthington-Wilmer J., Moritz C., Hall L. & Toop J. (1994) Extreme population structuring in the threatened ghost bat, *Macroderma gigas*: evidence from mitochondrial DNA. *Proceedings of the Royal Society of London B: Biological Sciences* 257(1349), 193-198.

Xu L., He C., Shen C., Jiang T., Shi L., Sun K., Berquist S. W. & Feng J. (2010) Phylogeography and population genetic structure of the great leaf-nosed bat (*Hipposideros armiger*) in China. *Journal of Heredity* 101(5), 562-572.

SUPPLEMENTARY MATERIAL

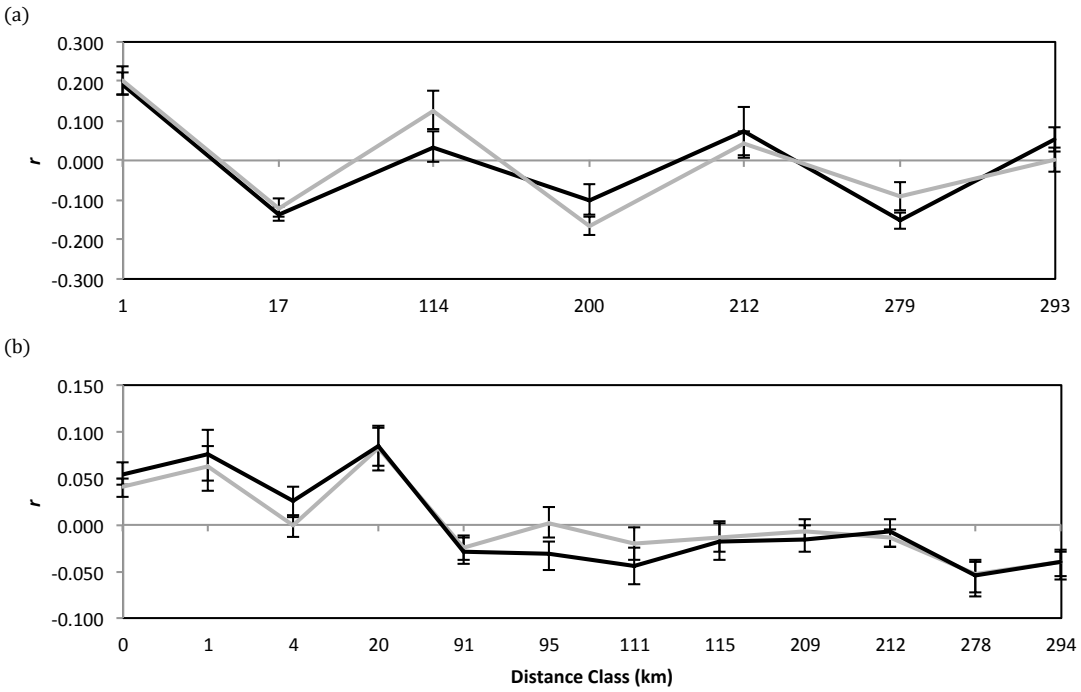


Figure S1. Spatial autocorrelogram showing pairwise relatedness between females (black) and males (grey) for a) *H.* sp. C and b) *H.* sp. D.

Acoustic analysis

Kruskal-Wallis tests for differences in call peak frequency by location:

Hipposideros sp. C: $\chi^2 = 23.21$, df = 6, $p = 0.0007$

Hipposideros sp. D: $\chi^2 = 6.97$, df = 7, $p = 0.43$

Table S1. *P*-values for pairwise Mann-Whitney-Wilcoxon tests for differences in peak frequency between locations for *H.* sp. C

	Akpafu	**Buoyem**	**Kwamang**
Buoyem	<0.000001	—	—
Kwamang	0.0066	1.0	—
Likpe	0.0007	1.0	1.0

Table S2. *P*-values for pairwise Mann-Whitney-Wilcoxon tests for differences in peak frequency between locations for *H.* sp. D

	Akpafu	**Buoyem**	**Kwamang**	
Buoyem	1.0	—	—	—
Forikrom	1.0	1.0	—	—
Kwamang	1.0	0.79	1.0	—
Likpe	1.0	1.0	1.0	0.63

CHAPTER 7
General discussion

Image: Luci Kirkpatrick

In this thesis I investigated ecology and CoV epidemiology in West African bats. My four main objectives were: 1) to investigate the prevalence of CoVs in Ghanaian cave-dwelling bats; 2) to elucidate risk factors in CoV infection in bats; 3) to resolve taxonomic uncertainty in CoV-hosts in *Hipposideros caffer* and *Hipposideros ruber*, and 4) to explore patterns of genetic and acoustic diversity in the *H. caffer* group in West Africa. The findings of this thesis add to knowledge of CoV ecology in bats, as well as to the evolution and patterns of diversity within a relatively widespread, but understudied, group of Afrotropical bats that are natural CoV hosts. These findings are used to address broader questions about zoonoses, wildlife management, and the evolution of cryptic species in bats.

The first part of my thesis focuses on bat CoV epidemiology and ecology. My research demonstrates that CoVs are widespread in cave-dwelling bats in Ghana (Chapters 2 and 3). Of a total of 17 bat species sampled in Ghana, we found CoVs in six species from two bat families. We further identified seasonal, demographic and ecological risk factors in CoV infection in bats.

In Chapter 2, we tested for novel betacoronaviruses related to the MERS-CoV (formerly hCoV-EMC/2012) in more than 4700 bats from 17 species sampled in Ghana, and an additional 272 bats from four species sampled in Europe. We found a novel relative of MERS-CoV in *Nycteris* cf. *gambiensis* (Nycteridae) in Ghana, and close relatives of a previously described bat CoV, VM314 (Reusken *et al.* 2010), in three species of bats in Europe, *Pipistrellus pipistrellus, P. nathusii,* and *P. pygmaeus* (Vespertilionidae). Previously, only verspertilionid bats were known to host MERS-CoV-related betacoronaviruses (Reusken *et al.* 2010; Woo *et al.* 2006). The finding of a 2c-CoV in *N. gambiensis* demonstrated that this CoV clade can infect two different host bat families. We report high prevalence of 2c-CoV in bats in Europe and Ghana. This discovery, together with findings of betacoronaviruses closely related to MERS-CoV in vespertilionid bats since then (De Benedictis *et al.* 2014; Ithete *et al.* 2013; Yang *et al.* 2014), and in guano from a cave housing multiple bat species (Wacharapluesadee *et al.* 2013), supports the hypothesis that bats may be the natural reservoir for MERS-CoV.

We showed preliminary evidence that CoV prevalence was significantly higher for juvenile and lactating female *N.* cf. *gambiensis* (Chapter 2). We further explored risk factors for infection with four bat CoVs in Chapter 3. In this longitudinal study, we tested more than 7000 bats from 17 species for four CoVs, over two years of regular sampling. We found CoVs in six bat species from two families (Hipposideridae, Nycteridae), in the genera *Hipposideros* and *Nycteris*. The viruses included a MERS-CoV relative (Chapter 2), two viruses in the SARS-like CoV clade, and one relative of the human common cold virus hCoV-229E (Pfefferle *et al.* 2009). We examined temporal patterns and ecological and demographic risk factors for CoV infection. Our findings show a strong seasonal effect on CoV infection rates, and a strong association with age, whereby juvenile bats are at greater risk of infection. Low body condition also conferred higher risk, a worrisome discovery since body condition is likely affected by the ongoing biodiversity loss due to anthropogenic impacts. These findings provide new insights into the ecological, demographic and temporal processes that influence CoV infection dynamics in bats, contributing new knowledge to how temporal and ecological factors impact upon wildlife disease patterns. They also carry real-world implications for public health management and the prevention of the transmission of these viruses from bats to domestic animals or humans. We suggested avenues on which to focus future strategies for the prediction and prevention of zoonotic CoV outbreak, including avoiding consumption of juvenile bats and seasonally avoiding direct and indirect contact with bats.

The second half of my thesis focussed on the ecology and genetic diversity of the *Hipposideros caffer* species complex. In Chapter 3, we found bats identified as *Hipposideros* cf. *ruber*, as well as *H. caffer tephrus*, infected with CoVs. *Hipposideros* cf. *ruber* was infected with three of the four CoVs we examined. Following from this finding, we aimed to resolve taxonomic uncertainty and to investigate patterns of genetic diversity and dispersal in this understudied group. To this end, in Chapter 4, I developed microsatellite markers to provide a tool for

investigating genetic diversity and structure in *Hipposideros caffer* and *H. ruber*, which I utilized in Chapters 5 and 6.

In Chapter 5 we examined cryptic diversity in the *H. caffer* complex using microsatellite markers (developed in Chapter 4), acoustic and morphometric data. We found that four previously described mitochondrial DNA clades of *H. caffer* and *H. ruber* represent distinct species, with no evidence of interbreeding between groups. We proposed the continued use of the *Hipposideros caffer tephrus* nomenclature, and the interim names *Hipposideros* species B, C, and D, until further research can elucidate the complicated systematics. Further investigation of the deep mitochondrial divergences within lineages throughout the Afrotropics, using multiple lines of evidence, would be beneficial for resolving the remaining taxonomic uncertainty in this group.

Further, we explored evolutionary processes that generate genetic and acoustic diversity in these bats. Bayesian divergence dating, based on mtDNA, indicated a Pliocene origin of these species, with intraspecific lineages corresponding to the Pleistocene. These findings are consistent with the species pump hypothesis, which emphasizes an important role of climatic fluctuations and in species richness of forest refugia. We examined hypotheses for acoustic divergence between species, with the data consistent with social selection, reproductive character displacement or drift, but ruling out a role for resource partitioning.

In the final data chapter, we characterised patterns of intraspecific genetic diversity in *Hipposideros* species B, C, and D. We found surprisingly high levels of genetic structure across a relatively fine scale in *H.* sp. C and D, and a pattern of isolation by distance for lineage D, suggesting limited dispersal ability. As *H.* sp. B was sampled at only three sites in close proximity, it was not possible to draw strong conclusions about their genetic connectivity on the same scale. We also found interesting differences in phylogeography and demographic history in these species. We used these data to explore the assumption, often invoked in wildlife management, that closely related, co-distributed taxa with high similarity in ecological and natural history traits will have highly similar responses to extrinsic forces. We found that despite similar ecological and

natural history traits, these species show surprisingly different responses to the same historical and contemporary processes. Our findings emphasize the need to be wary of inferring intraspecific genetic patterns from data collected from closely related species.

CONCLUSIONS

This project addressed an important knowledge gap in the disease ecology of CoVs in bats. It also aimed to address systematic, ecological and evolutionary questions related to the recognised species *Hipposideros caffer* and *H. ruber*, which host several CoVs in Ghana. The findings of this thesis have implications for the prediction and prevention of zoonotic disease and management or control strategies, as well as for the conservation of genetic variation and evolutionary processes.

Our findings on risk factors in CoV infection highlight behaviours and times of year that carry a greater risk of virus spillover to humans or domestic animals, from which we can infer direct recommendations. Further, this information, along with knowledge generated on genetic diversity, connectivity, and dispersal may be used to parameterize models to forecast risk for the prediction and prevention of outbreaks and spread of zoonotic disease. This project has aided taxonomic resolution within the *Hipposideros caffer* group, and identified morphological and acoustic measures to aid diagnosis in the field, which will allow greater accuracy in future surveillance and prevalence estimates of CoVs in these species.

The confirmation that *H. caffer* and *H. ruber* represent four separate species provides a foundation for further work aiming to reassess the conservation statuses of these species in West Africa and the broader Afrotropics. Our findings shed light on the evolution of genetic and acoustic diversity in this group, suggesting an influential role for climatic fluctuations in elevating the species richness of forest refugia. Finally, the contrasting patterns of intraspecific diversity and demographic history challenged the common assumption that ecologically similar species have similar responses to similar processes, emphasizing the need to be wary of inferring intraspecific genetic patterns on the basis of data from closely related or similar species.

The increased incidence of EIDs in recent decades is unlikely to wane without dramatic interventions into the proposed drivers, such as biodiversity loss, misuse of antibiotics, human-wildlife contact, and, increasingly and worryingly, anthropogenic climate change. Climate change is already being observed to influence disease patterns and emergence of diseases around the globe (Lafferty 2009; Naicker 2011; Rosenthal 2009). Studies such as this help to build a stronger knowledge base from which to tackle the growing threat, to both humans and wildlife, that is posed by emerging infectious diseases.

REFERENCES

De Benedictis P., Marciano S., Scaravelli D., Priori P., Zecchin B., Capua I., Monne I. & Cattoli G. (2014) Alpha and lineage C betaCoV infections in Italian bats. *Virus Genes* 48(2), 366-371.

Ithete N. L., Stoffberg S., Corman V. M., Cottontail V. M., Richards L. R., Schoeman M. C., Drosten C., Drexler J. F. & Preiser W. (2013) Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerging Infectious Diseases* 19(10), 1697-1699.

Lafferty K. D. (2009) The ecology of climate change and infectious diseases. *Ecology* 90(4), 888-900.

Naicker P. (2011) The impact of climate change and other factors on zoonotic diseases. *Archives of Clinical Microbiology* 2(4), 1-6.

Pfefferle S., Oppong S., Drexler J., Gloza-Rausch F., Ipsen A., Seebens A., Müller M., Annan A., Vallo P., Adu-Sarkodie Y., Kruppa T. & Drosten C. (2009) Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerging Infectious Diseases* 15(9), 1377-1384.

Reusken C. B. E. M., Lina P. H. C., Pielaat A., de Vries A., Dam-Deisz C., Adema J., Drexler J. F., Drosten C. & Kooi E. A. (2010) Circulation of group 2 coronaviruses in a bat species

common to urban areas in Western Europe. *Vector-Borne and Zoonotic Diseases* 10(8), 785-791.

Rosenthal J. (2009) Climate change and the geographic distribution of infectious diseases. *EcoHealth* 6(4), 489-495.

Wacharapluesadee S., Sintunawa C., Kaewpom T., Khongnomnan K., Olival K. J., Epstein J. H., Rodpan A., Sangsri P., Intarut N. & Chindamporn A. (2013) Group C betacoronavirus in bat guano fertilizer, Thailand. *Emerging Infectious Diseases* 19(8), 1349-1351.

Woo P., Lau S., Li K., Poon R., Wong B., Tsoi H., Yip B., Huang Y., Chan K. & Yuen K. (2006) Molecular diversity of coronaviruses in bats. *Virology* 351(1), 180-187.

Yang L., Wu Z., Ren X., Yang F., Zhang J., He G., Dong J., Sun L., Zhu Y. & Zhang S. (2014) MERS–related betacoronavirus in *Vespertilio superans* bats, China. *Emerging Infectious Diseases* 20(7), 1260-1261.

APPENDIX 1
Evidence for an ancestral association of human coronavirus 229E with bats

Image: Florian Gloza-Rausch

V. M. Corman, H. J. Baldwin, A. F. Tateno, R. M. Zerbinati, A. A. Annan, M. Owusu, E. E. Nkrumah, G. D. Maganga, S. K. Oppong, Y. Adu-Sarkodie, P. Vallo, L. V. R. F. da Silva Filho, E. M. Leroy, V. Thiel, L. van der Hoek, L. L. M. Poon, M. Tschapka, C. Drosten and J. F. Drexler. (2015) *Journal of Virology*, 89(23):11858-70.

ABSTRACT

We previously showed that close relatives of human coronavirus (HCoV)-229E exist in African bats. The small sample and limited genomic characterizations prevented further analyses so far. Here, we tested 2,087 faecal specimens from 11 bat species sampled in Ghana for HCoV-229E-related viruses by RT-PCR. Only hipposiderid bats tested positive. To compare the genetic diversity of bat viruses and HCoV-229E, we tested historical isolates and diagnostic specimens sampled globally over 10 years. Bat viruses were five- to sixfold more diversified than HCoV-229E in *RNA-dependent RNA polymerase* (*RdRp*) and *Spike* genes. In phylogenetic analyses, HCoV-229E strains were monophyletic and not intermixed with animal viruses. Bat viruses formed three large clades in close and more distant sister relationship. A recently described 229E-related alpaca virus occupied an intermediate phylogenetic position between bat and human viruses. According to taxonomic criteria, human, alpaca and bat viruses form a single CoV species showing evidence for multiple recombination events. HCoV-229E and the alpaca virus showed a major deletion in the *Spike* S1 region compared to all bat viruses. Analyses of four full genomes from 229E-related bat CoVs revealed an eighth open reading frame (*ORF8*) located at the genomic 3'-end. *ORF8* also existed in the 229E-related alpaca virus. Re-analysis of HCoV-229E sequences showed a conserved transcription regulatory sequence preceding remnants of this ORF, suggesting its loss after acquisition of a 229E-related CoV by humans. These data suggested an evolutionary origin of 229E-related CoVs in hipposiderid bats, hypothetically with camelids as intermediate hosts preceding the establishment of HCoV-229E.

One sentence summary: HCoV-229E-related bat coronaviruses are genetically highly diversified and suggest HCoV-229E acquired major genomic deletions upon host switching, potentially involving camelids as intermediate hosts.

Keywords: Africa, coronavirus, bats, camelids, HCoV-229E, zoonoses

Importance

The ancestral origins of major human coronaviruses (HCoV) likely involve bat hosts. Here, we provide conclusive genetic evidence for an evolutionary origin of the common cold virus HCoV-229E in hipposiderid bats by analyzing a large sample of African bats and characterizing several bat viruses on a full genome level. Our evolutionary analyses show that animal and human viruses are genetically closely related, can exchange genetic material and form a single viral species. We show that the putative host switches leading to the formation of HCoV-229E were accompanied by major genomic changes including deletions in the viral spike glycoprotein gene and loss of an open reading frame. We re-analyze a previously described genetically related alpaca virus and discuss the role of camelids as potential intermediate hosts between bat and human viruses. The evolutionary history of HCoV-229E likely shares important characteristics with that of the recently emerged highly pathogenic MERS-Coronavirus.

INTRODUCTION

Coronaviruses (CoV) are enveloped viruses with a single-stranded, positive-sense contiguous RNA genome of up to 32 kilobases. The subfamily *Coronavirinae* contains four genera termed *Alpha-*, *Beta-*, *Gamma-* and *Deltacoronavirus*. Mammals are predominantly infected by alpha- and betacoronaviruses, while gamma- and deltacoronaviruses mainly infect avian hosts (1, 2).

Four human coronaviruses (HCoVs) termed HCoV-229E, -NL63, -OC43 and -HKU1 circulate in the human population and mostly cause mild respiratory disease (3). HCoV-229E is frequently detected in up to 15% of specimens taken from individuals with respiratory disease (4-6). Although HCoV-229E can be detected in faecal specimens, HCoVs generally do not seem to play a role in acute gastroenteritis (7-9). Severe respiratory disease with high case-fatality rates is caused by severe acute respiratory syndrome (SARS)-CoV and Middle East respiratory syndrome (MERS)-CoV which emerged recently. HCoV-229E and HCoV-NL63 belong to the genus

Alphacoronavirus, while HCoV-OC43, HCoV-HKU1, SARS- and MERS-CoV belong to the genus *Betacoronavirus* (1, 10).

In analogy to major human pathogens including Ebola virus, rabies virus, mumps virus and hepatitis B and C viruses (11-16), the evolutionary origins of SARS- and MERS-CoV were traced back to bats (17-22). The genetic diversity of bat CoVs described over the last decade exceeds the diversity in other mammalian hosts (2). This has led to speculations on an evolutionary origin of all mammalian CoVs in bat hosts (23). Bats share important ecological features potentially facilitating virus maintenance and transmission, such as close contact within large social groups, longevity, and the ability of flight (13, 24).

How humans become exposed to remote wildlife viruses is not always clear (25). Human infection with SARS-CoV and MERS-CoV was likely mediated by peri-domestic animals. For SARS-CoV, the suspected source of infection were carnivores (26). Preliminary evidence suggested that these carnivore hosts may also have adapted SARS-CoV for human infection (27). For MERS-CoV, camelids are likely intermediate hosts, supported by circulation of MERS-CoV in camel herds globally and for prolonged periods of time (28-30). Whether MERS-CoV only recently acquired the capacity to infect humans is unclear.

The evolutionary origins of HCoV-229E are uncertain. In 2007, a syndrome of severe respiratory disease and sudden death was recognized in captive alpacas from the U.S. (31) and an alphacoronavirus genetically closely related to HCoV-229E was identified as the causative agent (32).

In 2009, we detected viruses in faecal specimens from 5 of 75 hipposiderid bats from Ghana and showed that these bat viruses were genetically related to HCoV-229E by characterizing their partial *RNA-dependent RNA polymerase* (*RdRp*) and *Nucleocapsid* genes (33). Lack of specimens containing high CoV RNA concentrations so far prevented a more comprehensive characterization of those bat viruses to further address their relatedness to HCoV-229E. Here, we tested more than 2,000 bats from Ghana for CoVs related to HCoV-229E. We

describe highly diversified bat viruses on a full genome level and analyze the evolutionary history of HCoV-229E and the genetically related alpaca CoV.

MATERIALS AND METHODS

Bat and human sampling

Bats were caught in the Ashanti region, central Ghana, during 2009-2011 as described previously (21). Archived anonymized respiratory specimens derived from patients sampled between 2002-2011 were obtained from Hong Kong/China, Germany, The Netherlands, Brazil and Ghana.

RNA purification, coronavirus detection and characterization

RNA was purified from approximately 20 mg of fecal material suspended in 500 µL RNAlater stabilizing solution using the MagNA Pure 96 system (Roche Penzberg, Germany). Elution volumes were 100 µL. Testing for CoV RNA was done using a real time RT-PCR assay designed to allow detection of HCoV-229E and all genetically related bat CoVs known from our pilot study (33). Oligonucleotide sequences were CoV229Elike-F13948m TCYAGAGAGGTKGTTGTTACWAAYCT, CoV229Elike-P13990m FAM (6-Carboxyfuorescein)-TGGCMACTTAATAAGTTTGGIAARGCYGG-BHQ1 (Black Hole Quencher 1) and CoV229Elike-R14138m CGYTCYTTRCCAGAWATGGCRTA. Testing used the SSIII RT-PCR Kit (Life Technologies, Karlsruhe, Germany) with the following cycling protocol in a LightCycler 480 (Roche, Penzberg, Germany): 20 min. at 50°C for reverse transcription, followed by 3 min. at 95°C and 45 cycles of 15 sec. at 95°C, 10 sec. at 58°C and 20 sec. at 72°C. CoV quantification relied on cRNA *in vitro* transcripts generated from TA-cloned peri-amplicons using the T7-driven Megascript (Life technologies, Heidelberg, Germany) kit as described previously (34). Partial *RdRp* gene sequences from real time RT-PCR-positive specimens were obtained as described previously (18). Full CoV genomes and *Spike* gene sequences were generated for those specimens containing highest CoV RNA concentrations using

sets of nested RT-PCR assays (primers available upon request) located along the HCoV-229E genome and designed to amplify small sequence islets. Sequence islets were connected by bridging long-range nested PCR using strain-specific primers (available upon request) and the Expand High Fidelity kit (Roche) on cDNA templates generated with the Superscript III reverse transcriptase (Life Technologies).

Phylogenetic analyses

Bayesian phylogenetic reconstructions were made using MRBAYES V3.1 (35) under assumption of a GTR+G+I nucleotide substitution model for partial *RdRp* sequences and the WAG amino acid substitution model for translated open reading frames (ORFs). Two million generations were sampled every 100 steps, corresponding to 20,000 trees of which 25% were discarded as burn-in before annotation using TREEANNOTATOR V1.5 and visualization using FIGTREE V1.4 from the BEAST package (36). Neighbour-joining phylogenetic reconstructions were made using MEGA5.2 (37) and a percentage nucleotide distance model, the complete deletion option and 1,000 bootstrap replicates. Genome comparisons were made using MEGA5.2 (37); SSE V1.1 (38) and recombination analyses were made using SIMPLOT V3.5 (39).

RESULTS

Specimens from 2,087 bats belonging to 11 species were available for PCR testing. Table 1 provides details on the overall sample composition and detection rates in individual bat species. Only bats belonging to the family Hipposideridae tested positive in 81 of 1,853 specimens (4.4%). All positive-testing bats had been morphologically identified in the field as either *Hipposideros* cf. *ruber* or *H. abae*. Those were the most abundant species within the sample. No HCoV-229E-related RNA was detected in the 17 available specimens from *H. jonesi* and *H.* cf. *gigas*.

Table 1. Overview of bats tested for 229E-related coronaviruses in Ghana

Species	*n*	Positives (%)
Coleura afra	68	0
Hipposideros abae	242	19 (7.8)
H. cf. *gigas*	12	0
H. jonesi	5	0
H. cf. *ruber*	1611	62 (3.8)
Nycteris cf. *gambiensis*	91	0
Rhinolophus alcyone	4	0
R. landeri	9	0
Taphozous perforatus	21	0
Lissonycteris angolensis	20	0
Rousettus aegyptiacus	4	0
Total	**2,087**	**81 (3.9)**

An 816 nucleotide (nt) fragment from the *RdRp* gene was obtained from 41 of the 81 positive specimens (GenBank accession nos. KT253259-KT253299). This fragment was used for further analysis as the 816 nt sequence yields improved resolution in inference of phylogeny as compared to shorter sequences derived from RT-PCR screening of field-derived samples (2). To expand the available genomic data for HCoV-229E, the 816 nt *RdRp* fragment was also sequenced from 23 HCoV-229E strains from patients sampled between 2002-2011 in China, Germany, The Netherlands, Brazil, and Ghana. In addition, the 816 nt *RdRp* fragment was sequenced from two historical HCoV-229E strains isolated in 1965 and the 1980ies (40) (GenBank accession nos. KT253300-KT253323). In analogy to the official taxonomic designation *SARS-related CoV* including human SARS-CoV and related CoVs from other animals (1), we hereafter restrict usage of the term HCoV-229E to the human virus and refer to the animal viruses as 229E-related CoV. Figure 1A shows a Bayesian phylogeny of the partial *RdRp* gene. The bat virus diversity we observed in our pilot study (represented by viruses Buoyem344 and Kwamang19) was expanded greatly. A phylogenetically basal virus termed Kwamang8 obtained within our pilot study was not detected again, although the present study contained specimens from the same cave and bat species. All human strains occupied an apical phylogenetic position and were not intermixed with any of the animal viruses. The recently described alpaca 229E-related CoV (32) clustered with two viruses obtained from hipposiderid bats in a parallel

Table 2. Coding capacity for the putative non-structural proteins of the novel bat 229E-related coronavirus

	KW2E-F151		F01A-F2		AT1A-F1		KW2E-F56	
	1st to last amino acid	Protein size	1st to last amino acid	Protein size	1st to last amino acid	Protein size	1st to last amino acid	Protein size
NSP1	Met^{1}-Gly^{111}	111	Met^{1}-Gly^{111}	111	Met^{1}-Gly^{111}	111	Met^{1}-Gly^{109}	109
NSP2	Asn^{112}-Gly^{897}	786	Asn^{112}-Gly^{897}	786	Asn^{112}-Gly^{897}	786	Asn^{110}-Gly^{895}	786
NSP3	Gly^{898}-Ala^{2494}	1597	Gly^{898}-Ala^{2494}	1597	Gly^{898}-Ala^{2492}	1595	Gly^{896}-Ala^{2489}	1594
NSP4	Gly^{2495}-Gln^{2975}	481	Gly^{2495}-Gln^{2975}	481	Gly^{2493}-Gln^{2973}	481	Gly^{2490}-Gln^{2970}	481
NSP5	Ala^{2976}-Gln^{3277}	302	Ala^{2976}-Gln^{3277}	302	Ala^{2974}-Gln^{3275}	302	Ala^{2971}-Gln^{3272}	302
NSP6	Ser^{3278}-Gln^{3556}	279	Ser^{3278}-Gln^{3556}	279	Ser^{3276}-Gln^{3553}	278	Ser^{3273}-Gln^{3551}	279
NSP7	Ser^{3557}-Gln^{3639}	83	Ser^{3557}-Gln^{3639}	83	Ser^{3554}-Gln^{3636}	83	Ser^{3552}-Gln^{3634}	83
NSP8	Ser^{3640}-Gln^{3834}	195	Ser^{3640}-Gln^{3834}	195	Ser^{3637}-Gln^{3831}	195	Ser^{3635}-Gln^{3829}	195
NSP9	Asn^{3835}-Gln^{3943}	109	Asn^{3835}-Gln^{3943}	109	Asn^{3832}-Gln^{3940}	109	Asn^{3830}-Gln^{3938}	109
NSP10	Ala^{3944}-Gln^{4078}	135	Ala^{3944}-Gln^{4078}	135	Ala^{3941}-Gln^{4075}	135	Ala^{3939}-Gln^{4073}	135
NSP11	Ser^{4079}-Glu^{4097}	19	Ser^{4079}-Glu^{4097}	19	Ser^{4076}-Glu^{4094}	19	Ser^{4074}-Glu^{4092}	19
NSP12	Ser^{4079}-Gln^{5005}	927	Ser^{4079}-Gln^{5005}	927	Ser^{4076}-Gln^{5002}	927	Ser^{4074}-Gln^{5000}	927
NSP13	Ala^{5006}-Gln^{5602}	597	Ala^{5006}-Gln^{5602}	597	Ala^{5003}-Gln^{5599}	597	Ala^{5001}-Gln^{5597}	597
NSP14	Ser^{5603}-Gln^{6120}	518	Ser^{5603}-Gln^{6120}	518	Ser^{5600}-Gln^{6117}	518	Ser^{5598}-Gln^{6115}	518
NSP15	Gly^{6121}-Gln^{6468}	348	Gly^{6121}-Gln^{6468}	348	Gly^{6118}-Gln^{6465}	348	Gly^{6116}-Gln^{6463}	348
NSP16	Ser^{6469}-Lys^{6768}	300	Ser^{6469}-Lys^{6768}	300	Ser^{6466}-Lys^{6766}	301	Ser^{6464}-Lys^{6763}	300

study from our groups in the Central African country Gabon (41). The two Gabonese bat-associated viruses differed from the alpaca 229E-related CoV by only 3.2% nucleotide content within the *RdRp* fragment. Hipposiderid bat CoVs were neither sorted by sampling sites, nor by their host species in their *RdRp* genes. Overall, bat 229E-related CoVs sampled over three years differed up to 13.5% in their nt and 3.3% in their amino acid (aa) sequences. Although the HCoV-229E dataset used for comparison was sampled over 50 years, the human-associated viruses showed 5-10fold less genetic diversity than bat viruses with only 1.4% nt and 0.7% aa variation. Because of the small sequence variation in HCoV-229E, Figure 1A contains only nine representative HCoV-229E strains. The neighbour-joining phylogeny shown in Figure 1B represents the high sequence identity between all HCoV-229E strains determined in this study.

To analyze to which extent bat 229E-related CoV show genetic variation, the *Spike* gene encoding the viral glycoprotein was characterized from 15 representative bat viruses (labelled with a triangle in Figure 1A). Figure 1C shows a Bayesian phylogenetic tree of the bat 229E-related CoV *Spike* gene sequences and HCoV-229E full *Spike* sequences sampled over 50 years. The bat viruses formed three genetically diverse lineage, of which two phylogenetically basal lineages contained bat viruses only. These lineages were sorted according to their sampling sites Kwamang (abbreviated KW) and Akpafu Todzi (abbreviated AT). A third lineage contained closely related bat viruses obtained from three different sample sites separated by several hundred kilometres (Buoyem, Kwamang and Forikrom) (21). These data suggested co-circulation of different *Spike* gene lineages within sampling sites as well as the existence of separate lineages between sites. However, the small number of viruses characterized from the phylogenetically basal bat clades 1 and 2 implies that caution should be taken in assertions on geographically separated *Spike* gene lineages. The alpaca 229E-related CoV and all HCoV-229E strains clustered in apical phylogenetic position compared to the bat viruses. The most closely related bat viruses from lineage 1 differed from HCoV-229E by 8.4-13.7%. The two other bat virus lineages were less related to HCoV-229E with 30.6-33.0% aa sequence distance.

Topologies of the Bayesian phylogenetic reconstructions of *RdRp* and *Spike* genes from bats and the alpaca were not congruent, compatible with past recombination events across animal 229E-related CoVs. To further investigate the genomic relationships of bat 229E-related CoVs and HCoV-229E, the full genomes from four representative bat viruses were determined directly from faecal specimens (labelled with circles in Figures 1A and C). Figure 2A shows that bat 229E-related CoV genomes comprise 28,014-28,748 nt, which exceeds the length of known HCoV-229E strains by 844-1,479 nt. As shown in Figure 2B, HCoV-229E and all bat viruses were closely related within the putative *ORF1ab*. This allowed the delineation of non-structural proteins (nsp) 1-16 for all bat viruses in analogy to HCoV-229E. Table 2 provides details on

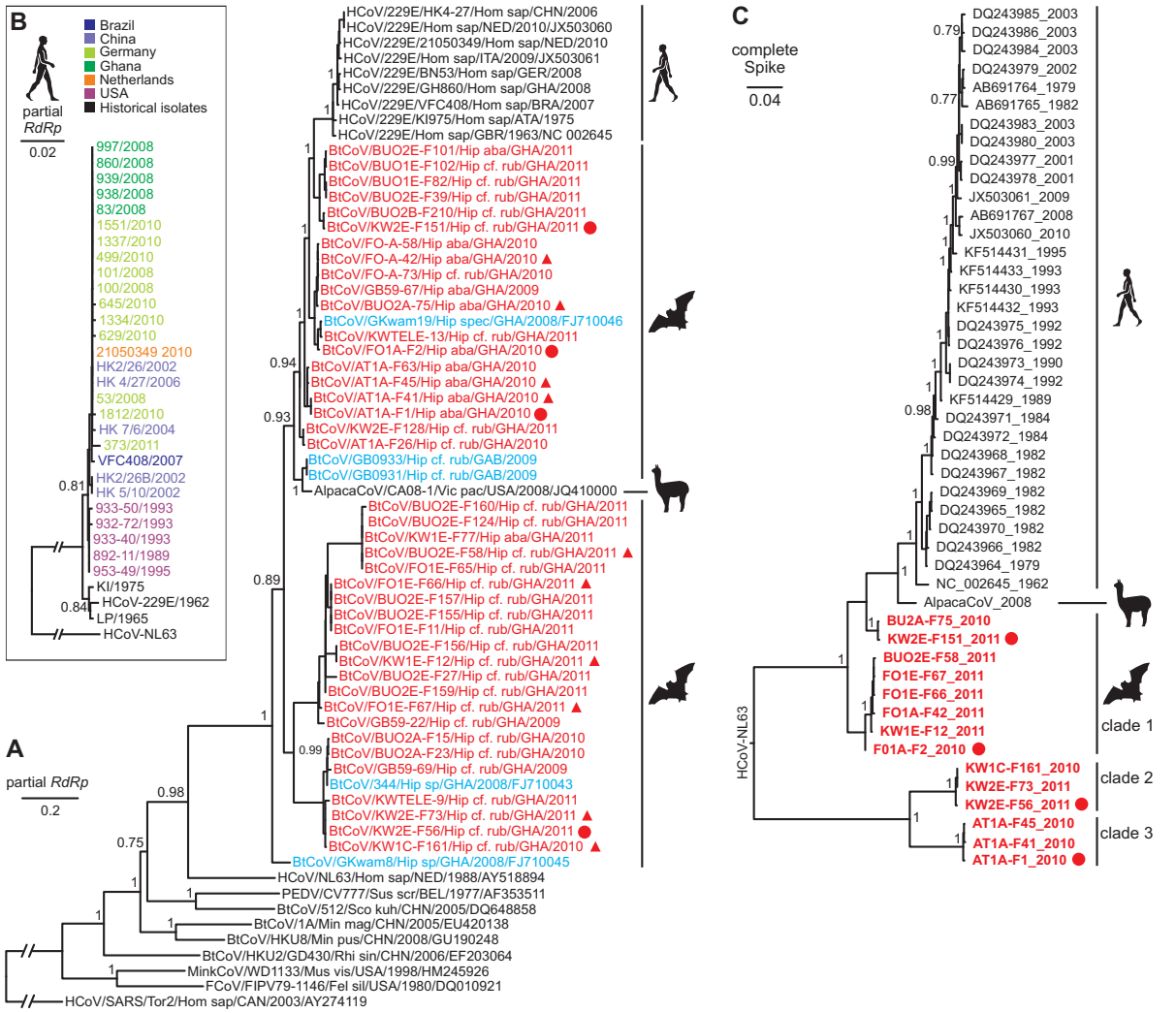


Figure 1. Phylogenetic relationships of the genus *Alphacoronavirus,* HCoV-229E strains and the novel bat viruses. A) Bayesian phylogeny of an 816 nucleotide *RdRp* gene sequence fragment corresponding to positions 13,891-14,705 in HCoV-229E prototype strain inf-1 (GenBank accession no. NC002645) using a GTR+G+I substitution model. SARS-coronavirus (CoV) was used as an outgroup. Viruses with additional sequence information generated in this study were marked with circles (full genome) or marked with triangles (*Spike* gene). Bat viruses detected in our previous studies from Ghana (Pfefferle *et al.* 2009) and Gabon are given in cyan (Maganga *et al.* 2014). B) Neighbour-joining phylogeny of the same *RdRp* gene fragment with a nucleotide percentage distance substitution model and the complete deletion option. The tree was rooted against HCoV-NL63. Viruses were coloured according to their origin. C) Bayesian phylogeny of the full *Spike* gene of bat 229E-related CoVs, the alpaca 229E-related CoV and HCoV-229E strains identified with GenBank accession numbers and year of isolation, using a WAG amino acid substitution model and HCoV-NL63 as an outgroup. The novel bat 229E–related CoVs are shown in boldface and red. Branches leading to the outgroup were truncated for graphical reasons as indicated by slashed lines. Values at nodes show support of grouping from posterior probabilities or 1,000 bootstrap replicates (only values above 0.7 were shown).

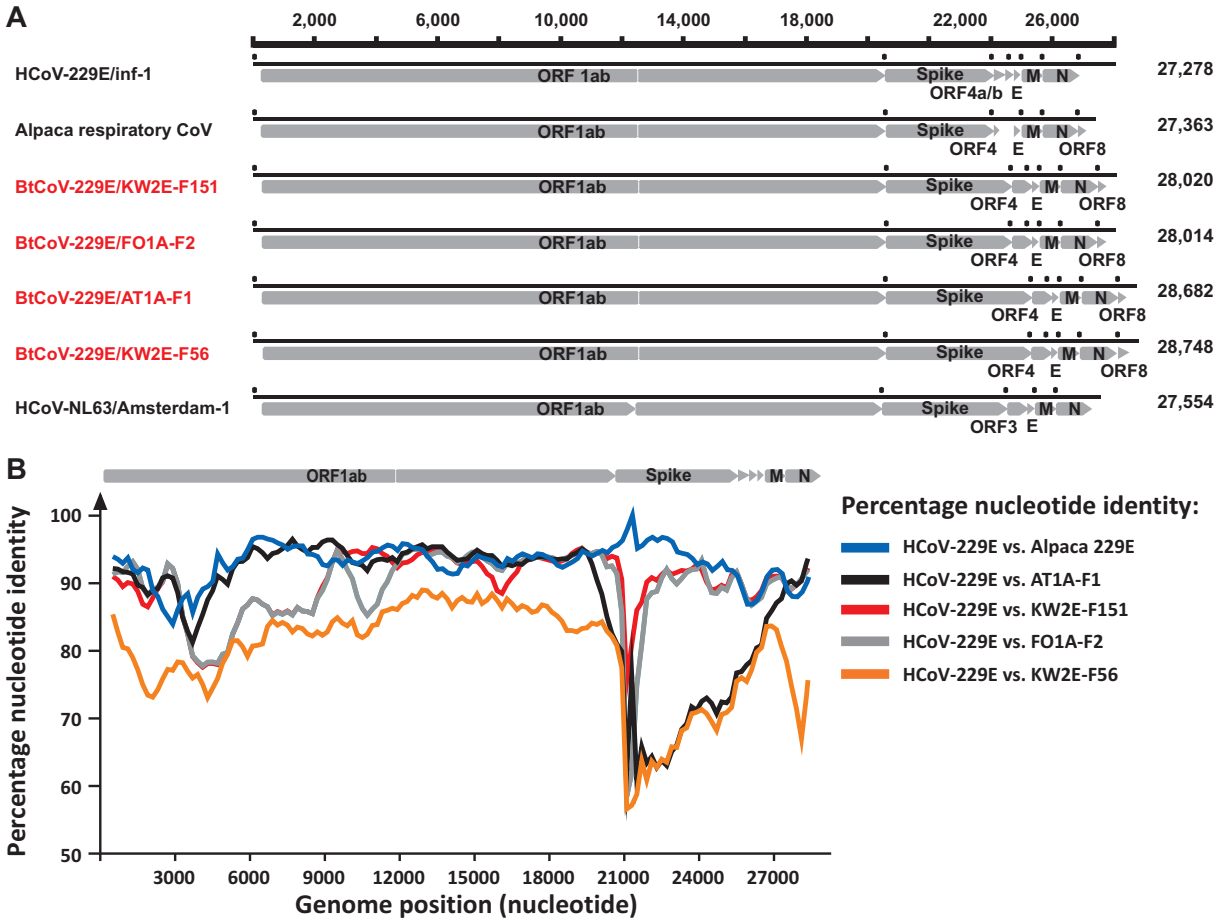


Figure 2. Genome organization of 229E-related coronaviruses and relationships between viruses from bats and humans. A) 229E-related CoV genomes represented by black lines; ORFs are indicated by grey arrows. Locations of transcription-regulatory core sequences (TRS) are marked by black dots. HCoV-NL63 is shown for comparison. B) Similarity plots generated using SSE V1.1 (38) using a sliding window of 400 and a step size of 40 nucleotides (nt). The HCoV-229E prototype strain inf-1 was used with animal viruses identified in the legend.

length and cleavage sites of the predicted nsp. Sequence identity in seven concatenated nsp is used by the International Committee for the Taxonomy of Viruses (ICTV) for CoV species designation (1). As shown in Table 3, the four fully sequenced bat viruses showed translated aa sequence identities of 93.3-97.1% with HCoV-229E. This was well above the 90% threshold established by the ICTV, indicating all bat 229E-related CoVs and HCoV-229E form a single species. Bat virus Kwamang8, which formed a phylogenetically basal sister-clade to the other bat

Table 3. Comparison of amino acid identities of seven conserved replicase domains of the bat 229E-related coronaviruses, HCoV-229E and the alpaca 229E-related coronavirus for species delineation

	Percentage amino acid sequence identity					
		Human Coronavirus 229E[a] vs.				
Domains	**within Bat 229E[b]**	**KW2E-F56**	**AT1A-F1**	**KW2-F151**	**F01A-F2**	**ACoV[c] vs Bat 229E[b]**
ADRP	75.6-100	75-75.6	91.1-92.9	84.5-85.1	84.5-85.1	76.8-90.5
NSP5(3CLpro)	90.7-100	90.4-90.7	97.4-97.7	96.4-96.7	97.4-97.7	90.4-97.4
NSP12 (RdRp)	97.5-100	95.7-96	97.3-97.6	96.9-97.3	97.2-97.7	97.3-98.9
NSP13 (NTPase/Hel)	97.2-100	96.5-97.2	97.2-97.8	97.3-98	98-98.7	97.8-99.3
NSP14 (ExoN/N7-MTase)	96.1-100	95-95.6	97.5-98.1	97.3-97.9	96.9-97.5	96.3-99.2
NSP15 (NendoU)	92.8-100	92.2	96.3-96.6	96.6-96.8	96.8-97.1	91.4-96.8
NSP16 (O-MT)	91.7-100	90.7-91	91.7-92	97.3-97	97.3-97.7	90.7 – 98.0
Concatenated domains	94.5-100	93.3-93.6	96.4-96.8	96.4-96.7	96.7-97.1	94.2-97.8

[a]including - HCoV 229E - Inf-1, HCoV 229E - 0349, HCoV 229E - J0304

[b]including - Bat CoV KW2E-F56, AT1A-F1, KW2E-F151, F01A-F2;

[c]ACoV - Alpaca Coronavirus

GenBank accession numbers of reference sequences: HCoV-229E - Inf-1: NC_002645.1; HCoV-229E - 0349: JX503060; HCoV-299E - J0304: JX503061; Alpaca CoV (ACoV): JQ410000

viruses and HCoV-229E, could not be sequenced on a full genome level. The aa sequence of the partial *RdRp* gene of Kwamang8 differed by only 3.3% from other bat viruses and HCoV-229E. Based upon previous comparisons of CoV *RdRp* sequences for tentative species delineation (2, 18), Kwamang8 forms part of the same species as the other bat viruses and HCoV-229E. This CoV species would also include the recently described alpaca 229E-related CoV (32), which showed 96.9-97.2% aa sequence identity with HCoV-229E and 94.2-97.8% with the bat viruses in the seven concatenated nsp domains.

Table 4. Amino acid identity between open reading frames of human, bat and camelid 229-related coronaviruses

	Percentage Amino Acid Sequence Identity						
	Human Coronavirus 229E[a] vs.						
	KW2E-F151	F01A-F2	AT1A-F1	KW2E-F56	AcoV	within Bat CoV[b]	ACoV[c] vs Bat CoV[b]
ORF1a	89.5 - 89.9	89.5 - 89.8	92.6 - 93.1	84.1 - 84.6	92.9 - 93.3	83.8 - 97.9	85.1 - 93.5
ORF1ab	92.5 - 92.9	92.6 - 93	94.2 - 94.6	88.3 - 88.8	94.6 - 95	88.7 - 98.3	89.3 - 95.2
Spike	87.5 - 91.6	87.4 - 91.4	67.2 - 68.9	67.2 - 69.1	92.8 - 94.4	66.8 - 92.4	69.7 - 90.8
ORF4	92.4 - 93.1	92.6 - 93.2	77.3 - 78.8	71.2 - 73.6	79.7 - 78.1	75.7 - 96.4	67.2 - 82.8
Envelope	89.6 - 90.9	89.6 - 90.9	77.6 - 78.9	78.7 - 80	89.6 - 90.9	77.3 - 98.7	77.3 - 100
Membrane	90.2 - 90.7	89.3 - 89.9	86.2 - 86.7	87.1 - 87.6	89.8 - 90.2	86.7 - 98.7	86.3 - 99.1
Nucleocapsid	90.7 - 92	90.2 - 91.5	88.6 - 90.4	75.8 - 76.6	88.4 - 89.7	78.7 - 99.5	78.2 - 94
ORFX/8	-	-	-	-	-	12.5 - 100	15.2 - 83.9

[a]including - HCoV 229E - Inf-1, HCoV 229E - 0349, HCoV 229E - J0304

[b]including - Bat CoV KW2E-F56, AT1A-F1, KW2E-F151, F01A-F2;

[c]ACoV - Alpaca Coronavirus

As shown in Figure 2A, all seven open reading frames (ORFs) known from HCoV-229E were found in bat 229E-related CoVs in the sequence *ORF1a/1b-Spike-ORF4-Envelope-Membrane-Nucleocapsid*. Amino acid identities between predicted ORFs of the bat viruses and HCoV-229E ranged from the 67.2-91.6% described above for the translated *Spike* genes to 88.3-94.6% (*ORF1ab*), with bat virus lineage 1 consistently showing highest aa sequence identities. Table 4 provides details for all sequence comparisons.

We looked for additional support for the existence of these predicted ORFs by analyzing the sequence context at their 5'-termini. This is because in CoVs, ORFs are typically preceded by highly conserved transcription regulatory sequence (TRS) elements (42). All putative ORFs from bat-229E related CoVs showed high conservation of the typical HCoV-229E TRS core sequence UCU C/A AACU and adjacent bases. Table 5 provides details on all putative TRS elements within bat 229E-related CoV genomes.

Table 5. Putative transcription regulatory sequences of the novel bat 229E-related coronaviruses and HCoV-229E

	HCoV-229E/ inf-1	KW2E-F151	F01A-F2	AT1A-F1	KW2E-F56
Leader	(62) UCUCAACUAAACN$_{220}$ (293) AUG	(62) UCUCAACUAAACN$_{220}$ (293) AUG	(62) UCUCAACUAAACN$_{220}$ (293) AUG	(62) UCUCAACUAAACN$_{220}$ (293) AUG	(62) UCUCAACUAAACN$_{220}$ (293) AUG
Spike	(20571) UCUCAACUAAAUAA A (20586) AUG	(20585) UCUCAACUAAAUAA A (20600) AUG	(20585) UCUCAACUAAAUAA A (20600) AUG	(20576) UCUCAACUAAAAA (20589) AUG	(20570) UCUCAACUAAGUA (20583) AUG
ORF4	(24054) UCAACUAAAN$_{38}$ (24101) AUG	(24644) UCAACUAAACN$_{38}$ (24691) AUG	(24638) UCAACUAAACN$_{38}$ (24685) AUG	(25290) UCAACUAAACN$_{38}$ (25337) AUG	(25258) UCAACUAAACN$_{38}$ (25304) AUG
Envelope	(24599) UCUCAACUAAN$_{152}$ (24762) AUG	(25190) UCUCAACUAACN$_{149}$ (25349) AUG	(25184) UCUCAACUAACN$_{149}$ (25343) AUG	(25836) UCUCAACUAACN$_{149}$ (25992) AUG	(25805) UCAACUAACN$_{131}$ (25962) AUG
Membrane	(24991) UCUAAACUAAACG ACA (25007) AUG	(25578) UCUAAACUAAACGA CA (25594) AUG	(25572) UCUAAACUAAACGA CA (25588) AUG	(26224) UCUAAACUAAACG (26237) AUG	(26185) UCUAAACUAAACG (26198) AUG
Nucleocapsid	(25680) UCUAAACUGAACGA AAAG (25698) AUG	(26270) UCUAAACUGAACGA AAAG (26288) AUG	(26264) UCUAAACUGAACGA AAAG (26282) AUG	(26934) UCUAAACUGAACGA AAACC (26953) AUG	(26874) UCUAAACUGAACGA AAACC (26893) AUG
ORF8		(27468) UCAACUAAAC (27478) AUG	(27462) UCAACUAAAC (27472) AUG	(28130) UCAACUAAAC (28141) AUG	(28124) UCAACUAAAC (28134) AUG

First bracket: Genome position of the first residue of the putative TRS sequence, second bracket: genome position of the first base of the start codon; $N_{lower\ case}$: number of base residues between end of the putative TRS sequence and start codon (where applicable)

Figure 3A shows Bayesian phylogenetic trees reconstructed for all individual ORFs. The alpaca 229E-related CoV clustered in intermediate position between HCoV-229E and the bat viruses in the *ORF1ab* and *Spike*, but with bat viruses only in *Membrane*, *Envelope*, *Nucleocapsid*, and *ORF4*. The divergent topologies again suggested recombination events in 229E-related CoVs. To find further evidence for recombination events and identify genomic breakpoints, 229E-related CoVs were analyzed by bootscanning. As shown in Figure 3B, bootscanning supported multiple recombination events involving HCoV-229E, bat 229E-related CoVs and the alpaca 229E-related CoV. Major recombination breakpoints occurred within the *ORF1ab* and the beginning of the *Spike* gene, compatible with previous analyses of CoV recombination patterns (2) and the divergent topologies between the *RdRp* and *Spike* genes noted above. Bootscanning also suggested a potential genomic breakpoint within the *Spike* gene, mapping to the borders of the S1 (associated with receptor binding) and S2 domains (associated with membrane fusion). This would be consistent with previous evidence supporting intra-*Spike* recombination events in bat-associated CoVs (43). To obtain further support for potential

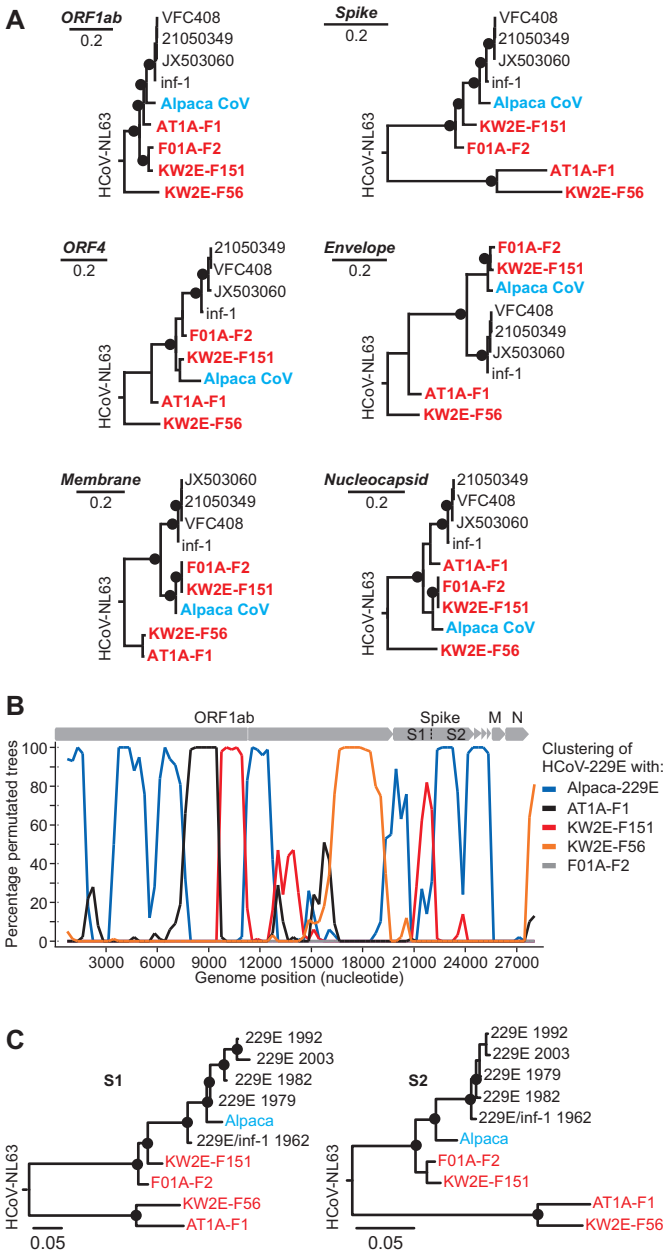


Figure 3. Bayesian phylogenies of major open reading frames and recombination analysis of HCoV-229E and related animal viruses. A) Phylogenies were calculated with a WAG amino acid substitution model. The novel bat viruses are shown in red. The alpaca CoV is shown in cyan. Filled circles, posterior probability support exceeding 0.95, scale bar corresponds to genetic distance. Details on the origin of HCoV-229E strain VFC408 which was generated for this study can be retrieved from (69). Branches leading the outgroup HCoV-NL63 were truncated for graphical reasons. B) Bootscan analysis using the Jukes-Cantor algorithm with a sliding window of 1,500 and a step size of 300 nt. The HCoV-220E inf-1 strain was used with animal 229E-related viruses as identified in the legend. C) Phylogenies of the S1 and S2 subunit were calculated according to A. One representative HCoV-229E strain was selected per decade according to (70); GenBank accession nos. DQ243974, DQ243964, DQ243984, DQ243967.

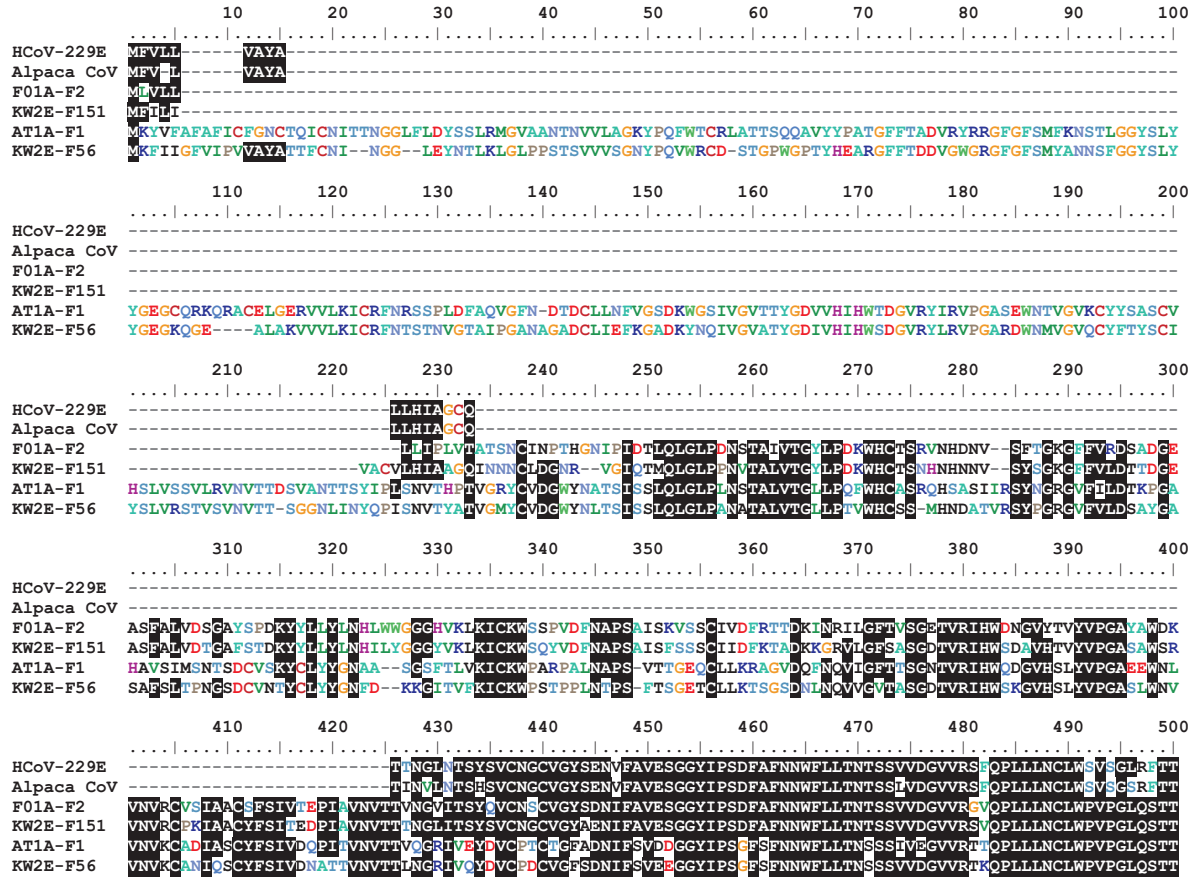


Figure 4. Amino acid sequence alignment of the 5'-end of the *Spike* gene of HCoV-229E and related animal viruses. Amino acid alignment of the first part of the *Spike* gene of 229E-related CoVs including four bat 229E-related CoVs, the alpaca 229E-related CoV and the HCoV-229E inf-1 strain. Conserved amino acid residues are marked in black, sequence gaps are represented by hyphens.

intra-*Spike* recombination events, separate phylogenetic reconstructions for the S1 and the S2 domains were made. As shown in Figure 3B, these Bayesian phylogenetic reconstructions supported recombination events involving the alpaca 229E-related CoV and HCoV-229E, but not the bat 22E-related CoVs. In the S1 domain, the alpaca 229E-related CoV clustered with clinical HCoV-229E strains, while the HCoV-229E reference strain inf-1 isolated in 1962 clustered in phylogenetically basal sister relationship. Only in the S2 domain, the intermediate position of the alpaca compared to bat and human 229E-related CoVs noted before in comparisons of the full *Spike* was maintained. These data may hint at recombination events

between HCoV-229E and the alpaca virus and further supported genetic compatibility between these two viruses belonging to one CoV species.

Three major differences existed between HCoV-229E, the alpaca 229E-related CoV and the bat 229E-related CoVs. The first of these differences occurred in the putative *ORF4*. Similar to HCoV-229E strains characterized from clinical specimens, a contiguous *ORF4* existed in all bat viruses that was 156-164 aa residues longer than the alpaca 229E-related CoV *ORF4*. Re-analysis of the putative *ORF4* sequence of the alpaca 229E-related CoV showed that this apparently shorter *ORF4* was due to an insertion of a single cytosine residue at position 181. Without this putative insertion, the alpaca 229E-related CoV *ORF4* showed the same length as homologous ORFs in bat 229E-related CoVs and HCoV-229E. Since the HCoV-229E *ORF4* is known to accumulate mutations in cell culture (40), the apparently truncated ORF in the alpaca 229E-related CoV isolate may thus not occur *in vivo*. The extended *ORF4* of the alpaca 229E-related CoV would be most closely related to bat viruses from clade 1 with 5.5% aa sequence distance, compared to at least 8.8% distance from HCoV-229E strains.

The second difference was a considerably longer S1 portion of the bat 229E-related CoV *Spike* genes compared to HCoV-229E. Figure 4 shows that the three bat lineages contained 185-404 additional aa residues upstream of the putative receptor binding domain (44, 45) compared to HCoV-229E. Bat lineage 1 which was phylogenetically most closely related to HCoV-229E carried the smallest number of additional aa residues. Of note, the alpaca 229E-related CoV was identical to HCoV-229E in the number of aa residues within this region of the *Spike* gene.

The third major difference was the existence of an additional putative ORF downstream of the *Nucleocapsid* gene in all bat viruses. Non-homologous ORFs of unknown function downstream the *Nucleocapsid* occur in several alpha- and betacoronaviruses, including feline infectious peritonitis virus (FIPV), transmissible gastroenteritis virus of swine (TGEV), *Rhinolophus* bat CoV HKU2, *Scotophilus* bat CoV 512, *Miniopterus* bat CoV HKU8 (23), the *Chaerephon* bat CoVs BtKY22/BtKY41, the *Cardioderma* bat CoV BtKY43 (46) and bat CoV

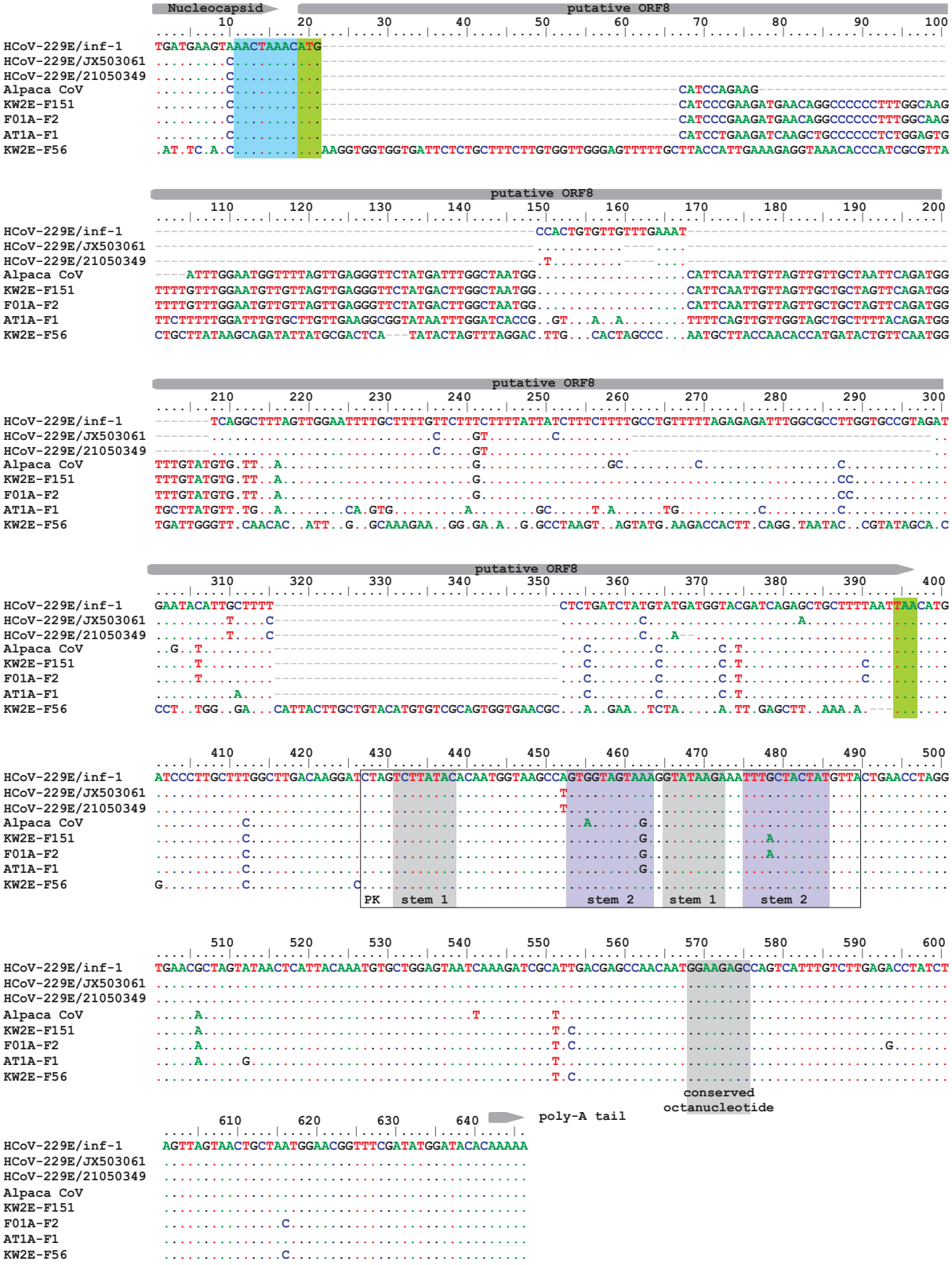


Figure 5. Nucleotide sequence alignment of the genomic 3'-end of HCoV-229E and related animal viruses. Nucleotide alignment of the genome region downstream the *Nucleocapsid* gene including four bat 229E-related CoV, the alpaca 229E-related CoV and representative HCoV-229E full genomes identified with GenBank accession number or strain name. Dots represent identical nucleotides, hyphens represent sequence gaps. Grey bars above

alignments indicate open reading frames and the beginning of the poly-A tail. The putative start and stop codon of *ORF8* is labelled lime green, the corresponding putative TRS element is marked blue. The conserved genomic sequence elements and the highly conserved stem elements forming part of the pseudo-knot (PK) were marked with grey and purple background.

HKU10 from Chinese *Hipposideros* and *Rousettus* species (47). In the genus *Betacoronavirus*, only *Bat CoV HKU9* from *Rousettus* and the genetically related *Eidolon* bat CoV BtKY24 (46) carry additional ORFs at this genomic position. No ORF in the 3'-terminal genome region is known from HCoV-229E. The alpaca 229E-related CoV contains an ORF at this position termed *ORFX* by Crossley *et al.* (32). In analogy to consecutive numbers used to identify HCoV-229E ORFs, we refer to this ORF as *ORF8* hereafter. The putative TRS context preceding *ORF8* was conserved in all bat 229E-related CoV and in the alpaca 229E-related CoV, suggesting that a corresponding subgenomic mRNA8 may exist. The 3'-UTR of bat 229E-related CoVs immediately followed the putative *ORF8*. This was supported by the existence of a conserved octanucleotide sequence and highly conserved stem elements forming part of the pseudo-knot typically located at the 5'-end of alphacoronavirus 3'-UTRs (48). As shown in Figure 5, HCoV-229E shows a high degree of sequence conservation compared to bat 229E-related CoVs and the alpaca 229E-related CoV in this genomic region, including a highly conserved putative TRS. Bioinformatic analyses (49-51) provided evidence for the presence of two transmembrane domains in the predicted proteins 8 of the alpaca and the genetically related bat 229E-related viruses. This may imply a role of the predicted protein 8 in coronaviral interactions with cellular or viral membranes.

As shown in Figure 5, one of the bat 229E-related CoV lineages represented by virus KW2E-F56 contained a highly divergent *ORF8*. In protein BLAST comparisons, the KW2E-F56 *ORF8* showed limited similarity to the putative *ORF7b* of HKU10 and to the putative *ORF8* located upstream of the *Nucleocapsid* of a Nigerian *Hipposideros* betacoronavirus termed Zaria

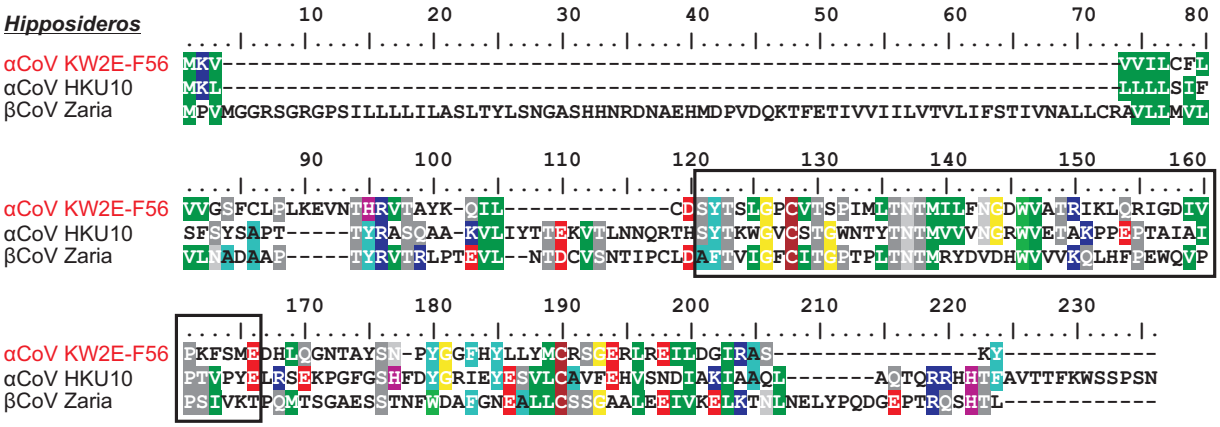


Figure 6. Amino acid sequence alignment of the putative *ORF8* from a bat 229E-related coronavirus and closest hits from two other hipposiderid bat coronaviruses. Conserved amino acid residues between sequence pairs are highlighted in colour according to amino acid properties, sequence gaps are represented by hyphens. The central domain showing higher sequence similarity between compared viruses is boxed for clarity. The 229E-related alphacoronavirus KW2E-F56 from a *Hipposideros* cf. *ruber* detected in this study is given in red, the alphacoronavirus HKU10 originated from a Chinese *H. pomona*, the betacoronavirus Zaria originated from a Nigerian *H. gigas*.

bat CoV (47, 52). This may hint at cross-genus recombination events between different hipposiderid bat CoVs in the past. However, overall aa sequence identity between these bat CoV ORFs was very low with maximally 28.2%. As shown in Figure 6, only the central part of these ORFs contained a stretch of 46 more conserved aa residues showing up to 39.1% sequence identity and 47.8% similarity (Blosum62 matrix). The origin and function of the divergent *ORF8* thus remain to be determined.

DISCUSSION

We characterize highly diverse bat CoVs on a full genome level and show that these viruses form one species together with HCoV-229E and a recently described virus from alpacas (32). We analyze the genomic differences between human, bat and alpaca 229E-related CoVs to elucidate potential host transitions during the formation of HCoV-229E.

A major difference between bat 229E-related CoVs and HCoV-229E was the *Spike* deletion in HCoV-229E compared to the bat viruses. Interestingly, the bat 229E-related CoV

lineage 1 which was phylogenetically most related to HCoV-229E also carried the smallest number of additional aa residues. Most chiropteran CoVs are restricted to the gastrointestinal tract, whereas HCoVs mainly replicate in the respiratory tract (2). The *Spike* deletion in HCoV-229E compared to ancestral bat viruses is thus noteworthy, since deletions in this protein have been associated with changes in coronaviral tissue tropism. This is best illustrated by TGEV, whose full-length Spike variants are associated with a dual tropism for respiratory and enteric tract, whereas the deleted variant termed porcine respiratory CoV (PRCV) mainly replicates in the respiratory tract (53). One could hypothesize that adaptation of bat 229E-related CoV lineage 1 to both non-chiropteran hosts and to respiratory transmission may have been easier compared to the other bat 229E-related CoV lineages.

Because the exact aa residues of the HCoV-229E RBD conveying cell entry are not known, it is difficult to predict whether the bat viruses may interact with the HCoV-229E cellular receptor Aminopeptidase N (45) or its *Hipposideros* homologue. Characterization of this bat molecule and identification of permissive cell culture systems may allow initial susceptibility experiments for chimeric viruses. Of note, although the alpaca 229E-related CoV was successfully isolated (32), no data on receptor usage and cellular tropism are available so far (2, 53).

Another major difference was the existence of an *ORF8* downstream the *Nucleocapsid* gene in bat 229E-related viruses and the detection of putative sequence remnants of this ORF in HCoV-229E. Hypothetically, deterioration of *ORF8* in HCoV-229E could have occurred due to loss of gene function in human hosts after zoonotic transmission from bats or intermediate hosts. This may parallel gradual deletions in the SARS-CoV accessory *ORF8* during the human epidemic compared to bat SARS-related CoVs (54) and is consistent with characterizations of HCoV-229E clinical strains showing high variability of this genomic region (55).

The virus-host association between 229E-related CoVs and the bat genus *Hipposideros* is strengthened by our virus detections in *Hipposideros* species in Ghana and in Gabon (41), which is separated from Ghana by about 1,800 km. The observed link between 229E-related

alphacoronaviruses and hipposiderid bats is paralleled by the detections of genetically closely related betacoronaviruses in different *Hipposideros* species from Ghana, Nigeria, Thailand and Gabon (33, 41, 52, 56), suggesting restriction of these CoVs to hipposiderid bat genera. Due to their proofreading capacity, CoVs show evolutionary rates of 10E-5 to 10E-6 substitutions per site per replication cycle, which is much slower than rates observed for other RNA viruses (57, 58). Our data thus suggest a long evolutionary history of 229E-related CoVs in Old World hipposiderid bats that greatly exceeds that of HCoV-229E in humans, confirming previous hypotheses from our group (33).

The putative role of the alpaca 229E-related CoV in the formation of HCoV-229E is unclear. Our data enable new insights into the evolutionary history of HCoV-229E. First, the alpaca 229E-related CoV contained an intact *ORF8* which was genetically related to the homologous gene in bat 229E-related CoVs. Second, genes of the alpaca CoV clustered either with bat viruses only or in intermediate position between bat viruses and HCoV-229E. Because the alpaca 229E-related CoV showed the same deletion in its *Spike* gene as HCoV-229E compared to bat 229E-related CoVs, it may be possible that alpacas represent a first host switch from bats followed by a second inter-host transfer from alpacas to humans. The relatedness of the alpaca 229E-related CoV to older HCoV-229E strains rather than to contemporary ones reported by Crossley *et al.* would be compatible with this scenario (32). However, the alpaca 229E-related CoV was reported only from captive animals in the U.S. and whether this virus is indeed endemic in New World alpacas is unclear. Additionally, the apparent intra-*Spike* recombination event may speak against a role of the alpaca virus as the direct ancestor of HCoV-229E. Further analyses will be required to confirm this putative recombination event, ideally including additional sequence information from old HCoV-229E strains. Furthermore, a hypothetical direct transfer of Old World bat viruses to New World alpacas appears geographically unfeasible. It would be highly relevant to investigate Old World camelids for 229E-related CoVs that may have been passed on to captive alpacas and that may represent direct ancestors of HCoV-229E.

Additional constraints to consider in the hypothetical role of camelids for the evolutionary history of 229E-related CoVs is the time and place of putative host switches from bats. Camels were likely introduced to Africa not earlier than 5,000 years ago from the Arabian Peninsula (59, 60) and could not possibly come into direct contact with West African *H.* cf. *ruber* or *H. abae* of the Guinean savanna. The majority of CoV species seems to be confined to host genera (2). Therefore, it may be possible that 229E-related CoV transmission was mediated through closely related species like *H. tephrus*, which occurs in the Sahel zone and comes into contact to populations of *H.* cf. *ruber* distantly related to those from the Guinean savanna (61). This bat species should be analyzed for 229E-related CoVs together with other genera of the family Hipposideridae, like *Asellia* or *Triaenops*, which are desert-adapted bats sharing their habitat with camelids both in Arabia and Africa and may harbor genetically related CoVs. An important parallel to this evolutionary scenario is the role of camelids for the emerging MERS-CoV (30, 62), whose likely ancestors also occur in bats (20, 21). However, we cannot rule out that the alpaca 229E-related CoV and HCoV-229E represent two independent zoonotic acquisitions from 229E-related CoVs existing in hipposiderid bats and potentially yet unknown intermediate hosts.

The existence of different serotypes in the expanded 229E-related CoV species is unclear. CoV neutralization is mainly determined by antibodies against the S protein, and particularly the S1 domain (63). The phylogenetic relatedness of the S1 domains from the alpaca 229E-related CoV and HCoV-229E suggests that these viruses form one serotype. The most closely related bat 229E-related CoV lineage showed 8.4% aa sequence distance in the translated *Spike* gene from HCoV-229E. This was comparable to the 7.8-18.6% aa distance between FIPV, TGEV und canine CoV, which belong to one CoV species (*Alphacoronavirus 1*) and for which cross-neutralization was observed (64). The about 30% Spike aa sequence distance between the other bat 229E-related lineages and HCoV-229E were comparable to the distance between HCoV-NL63 and HCoV-229E, which form two different serotypes (65). HCoV-229E thus likely forms

one serotype that includes the alpaca 229E- and potentially the most closely related bat 229E-related lineage, while the other bat 229E-related lineages may form different serotypes. In our study, lack of bat sera and absence of bat 229E-related CoV isolates prevented serological investigations. The generation of pseudotyped viruses carrying bat 229E-related *Spike* motifs may allow future serological studies. Of note, our joint analyses of Ghanaian patients with respiratory disease in this study and previous work from our group investigating Ghanaian villagers (66) showed that Ghanaians were infected with the globally circulating HCoV-229E, whereas no evidence of bat 229E-related CoV infecting humans was found. If serotypes existed in 229E-related CoVs, serologic studies may thus aid to elucidate putative exposure of humans and potential camelid intermediate hosts to these bat viruses.

It should be noted that throughout Africa, bats are consumed as wild game (67) and humans frequently live in close proximity of bat caves (68), including usage of bat guano as fertilizer and drinking water from these caves (21). These settings potentially facilitate the exposure of humans and their peri-domestic animals, including camelids, to these previously remote bat viruses.

In summary, HCoV-229E may be a paradigmatic example of the successful introduction of a bat CoV into the human population, possibly with camelids as intermediate hosts.

ACKNOWLEDGEMENTS

We thank Monika Eschbach-Bludau, Sebastian Brünink, Tobias Bleicker, Fabian Ebach and Thierno Diawo Dallo at the Institute of Virology, Bonn, for technical assistance and Ebenezer Kofi Badu, Priscilla Anti, Olivia Agbenyega, Florian Gloza-Rausch, Stefan Klose and Thomas Kruppa for their help during the organization and conducting of field work. For all capturing, sampling, and exportation of bat specimens, we obtained permission from the respective countries' authorities. This study was supported by the European Union FP7 projects EMPERIE

(contract number 223498) and ANTIGONE (contract number 278976) and by the German Research Foundation (DFG grants DR 772/3-1, KA1241/18-1 and DR 772/7-1, TH 1420/1-1).

REFERENCES

1. de Groot RJ, Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A.E., Holmes, K.V., Perlman, S., Poon, L., Rottier, P.J.M.,, Talbot PJ, Woo, P.C.Y. and Ziebuhr, J. 2012. Family *Coronaviridae*, p 806-820. *In* King A, Adams M, Carstens EB, Lefkowitz EJ (ed), Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses. Elsevier/Academic Press, Amsterdam; Boston.
2. Drexler JF, Corman VM, Drosten C. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral Res 101:45-56.
3. van der Hoek L. 2007. Human coronaviruses: what do they cause? Antivir Ther 12:651-658.
4. Walsh EE, Shin JH, Falsey AR. 2013. Clinical impact of human coronaviruses 229E and OC43 infection in diverse adult populations. J Infect Dis 208:1634-1642.
5. Berkley JA, Munywoki P, Ngama M, Kazungu S, Abwao J, Bett A, Lassauniere R, Kresfelder T, Cane PA, Venter M, Scott JA, Nokes DJ. 2010. Viral etiology of severe pneumonia among Kenyan infants and children. JAMA 303:2051-2057.
6. Gaunt ER, Hardie A, Claas EC, Simmonds P, Templeton KE. 2010. Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. J Clin Microbiol 48:2940-2947.
7. Esper F, Ou Z, Huang YT. 2010. Human coronaviruses are uncommon in patients with gastrointestinal illness. J Clin Virol 48:131-133.

8. Paloniemi M, Lappalainen S, Vesikari T. 2015. Commonly circulating human coronaviruses do not have a significant role in the etiology of gastrointestinal infections in hospitalized children. J Clin Virol 62:114-117.

9. Risku M, Lappalainen S, Rasanen S, Vesikari T. 2010. Detection of human coronaviruses in children with acute gastroenteritis. J Clin Virol 48:27-30.

10. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish Z, Perlman S, Poon LL, Snijder EJ, Stephens GM, Woo PC, Zaki AM, Zambon M, Ziebuhr J. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV); announcement of the Coronavirus Study Group. J Virol 87:7790-7792.

11. Drexler JF, Geipel A, Konig A, Corman VM, van Riel D, Leijten LM, Bremer CM, Rasche A, Cottontail VM, Maganga GD, Schlegel M, Muller MA, Adam A, Klose SM, Carneiro AJ, Stocker A, Franke CR, Gloza-Rausch F, Geyer J, Annan A, Adu-Sarkodie Y, Oppong S, Binger T, Vallo P, Tschapka M, Ulrich RG, Gerlich WH, Leroy E, Kuiken T, Glebe D, Drosten C. 2013. Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. Proc Natl Acad Sci U S A 110:16151-16156.

12. Drexler JF, Corman VM, Muller MA, Lukashev AN, Gmyl A, Coutard B, Adam A, Ritz D, Leijten LM, van Riel D, Kallies R, Klose SM, Gloza-Rausch F, Binger T, Annan A, Adu-Sarkodie Y, Oppong S, Bourgarel M, Rupp D, Hoffmann B, Schlegel M, Kummerer BM, Kruger DH, Schmidt-Chanasit J, Setien AA, Cottontail VM, Hemachudha T, Wacharapluesadee S, Osterrieder K, Bartenschlager R, Matthee S, Beer M, Kuiken T, Reusken C, Leroy EM, Ulrich RG, Drosten C. 2013. Evidence for novel hepaciviruses in rodents. PLoS Pathog 9:e1003438.

13. Drexler JF, Corman VM, Muller MA, Maganga GD, Vallo P, Binger T, Gloza-Rausch F, Rasche A, Yordanov S, Seebens A, Oppong S, Adu Sarkodie Y, Pongombo C, Lukashev AN, Schmidt-Chanasit J, Stocker A, Carneiro AJ, Erbar S, Maisner A, Fronhoffs F,

Buettner R, Kalko EK, Kruppa T, Franke CR, Kallies R, Yandoko ER, Herrler G, Reusken C, Hassanin A, Kruger DH, Matthee S, Ulrich RG, Leroy EM, Drosten C. 2012. Bats host major mammalian paramyxoviruses. Nat Commun 3:796.

14. Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Delicat A, Paweska JT, Gonzalez JP, Swanepoel R. 2005. Fruit bats as reservoirs of Ebola virus. Nature 438:575-576.

15. Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. 2006. Bats: important reservoir hosts of emerging viruses. Clin Microbiol Rev 19:531-545.

16. Quan PL, Firth C, Conte JM, Williams SH, Zambrana-Torrelio CM, Anthony SJ, Ellison JA, Gilbert AT, Kuzmin IV, Niezgoda M, Osinubi MO, Recuenco S, Markotter W, Breiman RF, Kalemba L, Malekani J, Lindblade KA, Rostal MK, Ojeda-Flores R, Suzan G, Davis LB, Blau DM, Ogunkoya AB, Alvarez Castillo DA, Moran D, Ngam S, Akaibe D, Agwanda B, Briese T, Epstein JH, Daszak P, Rupprecht CE, Holmes EC, Lipkin WI. 2013. Bats are a major natural reservoir for hepaciviruses and pegiviruses. Proc Natl Acad Sci U S A 110:8194-8199.

17. Wacharapluesadee S, Lumlertdacha B, Boongird K, Wanghongsa S, Chanhome L, Rollin P, Stockton P, Rupprecht CE, Ksiazek TG, Hemachudha T. 2005. Bat Nipah virus, Thailand. Emerg Infect Dis 11:1949-1951.

18. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, Seebens A, Niedrig M, Pfefferle S, Yordanov S, Zhelyazkov L, Hermanns U, Vallo P, Lukashev A, Muller MA, Deng H, Herrler G, Drosten C. 2010. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial *RNA-dependent RNA polymerase* gene sequences. J Virol 84:11336-11349.

19. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Crameri G, Zhang SY, Wang LF,

Daszak P, Shi ZL. 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535-538.

20. Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, Schoeman MC, Drosten C, Drexler JF, Preiser W. 2013. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. Emerg Infect Dis 19:1697-1699.

21. Annan A, Baldwin HJ, Corman VM, Klose SM, Owusu M, Nkrumah EE, Badu EK, Anti P, Agbenyega O, Meyer B, Oppong S, Sarkodie YA, Kalko EK, Lina PH, Godlevska EV, Reusken C, Seebens A, Gloza-Rausch F, Vallo P, Tschapka M, Drosten C, Drexler JF. 2013. Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. Emerg Infect Dis 19:456-459.

22. Woo PC, Wang M, Lau SK, Xu H, Poon RW, Guo R, Wong BH, Gao K, Tsoi HW, Huang Y, Li KS, Lam CS, Chan KH, Zheng BJ, Yuen KY. 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. J Virol 81:1574-1585.

23. Woo PC, Lau SK, Huang Y, Yuen KY. 2009. Coronavirus diversity, phylogeny and interspecies jumping. Exp Biol Med (Maywood) 234:1117-1127.

24. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, Mills JN, Timonin ME, Willis CKR, Cunningham AA, Fooks AR, Rupprecht CE, Wood JLN, Webb CT. 2013. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proceedings of the Royal Society B: Biological Sciences 280.

25. Wolfe ND, Dunavan CP, Diamond J. 2007. Origins of major human infectious diseases. Nature 447:279-283.

26. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302:276-278.

27. Graham RL, Baric RS. 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J Virol 84:3134-3146.

28. Chu D, Poon L, Gomaa MM, Shehata MM, Perera R, Abu Zeid D, El Rifay AS, Siu LY, Guan Y, Webby RW, Ali MA, Peiris M, Kayali G. 2014. MERS Coronaviruses in dromedary camels, Egypt. Emerg Infect Dis 20.

29. Muller MA, Corman VM, Jores J, Meyer B, Younan M, Liljander A, Bosch BJ, Lattwein E, Hilali M, Musa BE, Bornstein S, Drosten C. 2014. MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983-1997. Emerg Infect Dis 20:2093-2095.

30. Reusken CB, Haagmans BL, Muller MA, Gutierrez C, Godeke GJ, Meyer B, Muth D, Raj VS, Smits-De Vries L, Corman VM, Drexler JF, Smits SL, El Tahir YE, De Sousa R, van Beek J, Nowotny N, van Maanen K, Hidalgo-Hermoso E, Bosch BJ, Rottier P, Osterhaus A, Gortazar-Schmidt C, Drosten C, Koopmans MP. 2013. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infect Dis 13:859-866.

31. Crossley BM, Barr BC, Magdesian KG, Ing M, Mora D, Jensen D, Loretti AP, McConnell T, Mock R. 2010. Identification of a novel coronavirus possibly associated with acute respiratory syndrome in alpacas (*Vicugna pacos*) in California, 2007. J Vet Diagn Invest 22:94-97.

32. Crossley BM, Mock RE, Callison SA, Hietala SK. 2012. Identification and characterization of a novel alpaca respiratory coronavirus most closely related to the human coronavirus 229E. Viruses 4:3689-3700.

33. Pfefferle S, Oppong S, Drexler JF, Gloza-Rausch F, Ipsen A, Seebens A, Muller MA, Annan A, Vallo P, Adu-Sarkodie Y, Kruppa TF, Drosten C. 2009. Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. Emerg Infect Dis 15:1377-1384.

34. Drexler JF, Kupfer B, Petersen N, Grotto RM, Rodrigues SM, Grywna K, Panning M, Annan A, Silva GF, Douglas J, Koay ES, Smuts H, Netto EM, Simmonds P, Pardini MI,

Roth WK, Drosten C. 2009. A novel diagnostic target in the hepatitis C virus genome. PLoS Med 6:e31.

35. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

36. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969-1973.

37. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739.

38. Simmonds P. 2012. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res Notes 5:50.

39. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73:152-160.

40. Dijkman R, Jebbink MF, Wilbrink B, Pyrc K, Zaaijer HL, Minor PD, Franklin S, Berkhout B, Thiel V, van der Hoek L. 2006. Human coronavirus 229E encodes a single ORF4 protein between the spike and the envelope genes. Virol J 3:106.

41. Maganga GD, Bourgarel M, Vallo P, Dallo TD, Ngoagouni C, Drexler JF, Drosten C, Nakoune ER, Leroy EM, Morand S. 2014. Bat distribution size or shape as determinant of viral richness in african bats. PLoS One 9:e100172.

42. Thiel V, Herold J, Schelle B, Siddell SG. 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. J Gen Virol 82:1273-1281.

43. Corman VM, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, Drexler JF. 2014. Rooting the phylogenetic tree of Middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. J Virol 88:11297-11303.

44. Breslin JJ, Mork I, Smith MK, Vogel LK, Hemmila EM, Bonavia A, Talbot PJ, Sjostrom H, Noren O, Holmes KV. 2003. Human coronavirus 229E: receptor binding domain and neutralization by soluble receptor at 37 degrees C. J Virol 77:4435-4438.

45. Hofmann H, Simmons G, Rennekamp AJ, Chaipan C, Gramberg T, Heck E, Geier M, Wegele A, Marzi A, Bates P, Pohlmann S. 2006. Highly conserved regions within the spike proteins of human coronaviruses 229E and NL63 determine recognition of their respective cellular receptors. J Virol 80:8639-8652.

46. Tao Y, Tang K, Shi M, Conrardy C, Li KS, Lau SK, Anderson LJ, Tong S. 2012. Genomic characterization of seven distinct bat coronaviruses in Kenya. Virus Res 167:67-73.

47. Lau SK, Li KS, Tsang AK, Shek CT, Wang M, Choi GK, Guo R, Wong BH, Poon RW, Lam CS, Wang SY, Fan RY, Chan KH, Zheng BJ, Woo PC, Yuen KY. 2012. Recent transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault's rousettes to pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders. J Virol 86:11906-11918.

48. Williams GD, Chang RY, Brian DA. 1999. A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. J Virol 73:8349-8355.

49. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Honigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B. 2014. PredictProtein -- an open resource for online prediction of protein structural and functional features. Nucleic Acids Res 42:W337-343.

50. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567-580.

51. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845-858.

52. Quan PL, Firth C, Street C, Henriquez JA, Petrosov A, Tashmukhamedova A, Hutchison SK, Egholm M, Osinubi MO, Niezgoda M, Ogunkoya AB, Briese T, Rupprecht CE, Lipkin WI. 2010. Identification of a severe acute respiratory syndrome coronavirus-like virus in a leaf-nosed bat in Nigeria. MBio 1:e00208-10.

53. Sanchez CM, Gebauer F, Sune C, Mendez A, Dopazo J, Enjuanes L. 1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. Virology 190:92-105.

54. Anonymous. 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303:1666-1669.

55. Farsani SM, Dijkman R, Jebbink MF, Goossens H, Ieven M, Deijs M, Molenkamp R, van der Hoek L. 2012. The first complete genome sequences of clinical isolates of human coronavirus 229E. Virus Genes 45:433-439.

56. Gouilh MA, Puechmaille SJ, Gonzalez JP, Teeling E, Kittayapong P, Manuguerra JC. 2011. SARS-Coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. Infect Genet Evol 11:1690-1702.

57. Eckerle LD, Becker MM, Halpin RA, Li K, Venter E, Lu X, Scherbakova S, Graham RL, Baric RS, Stockwell TB, Spiro DJ, Denison MR. 2010. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. PLoS Pathog 6:e1000896.

58. Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR. 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. J Virol 81:12135-12144.

59. Mburu DN, Ochieng JW, Kuria SG, Jianlin H, Kaufmann B, Rege JE, Hanotte O. 2003. Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. Anim Genet 34:26-32.

60. Epstein H, Mason IL. 1971. The origin of the domestic animals of Africa. Africana Pub. Corp.

61. Vallo P, Guillén-Servent A, Benda P, Pires DB, Koubek P. 2008. Variation of mitochondrial DNA in the *Hipposideros caffer* complex (Chiroptera: Hipposideridae) and its taxonomic implications. Acta Chiropterologica 10:193-206.

62. Perera RA, Wang P, Gomaa MR, El-Shesheny R, Kandeil A, Bagato O, Siu LY, Shehata MM, Kayed AS, Moatasim Y, Li M, Poon LL, Guan Y, Webby RJ, Ali MA, Peiris JS, Kayali G. 2013. Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. Eurosurveillance 18:8-14.

63. Meyer B, Drosten C, Muller MA. 2014. Serological assays for emerging coronaviruses: challenges and pitfalls. Virus Res 194:175-183.

64. Horzinek MC, Lutz H, Pedersen NC. 1982. Antigenic relationships among homologous structural polypeptides of porcine, feline, and canine coronaviruses. Infect Immun 37:1148-1155.

65. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pohlmann S. 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc Natl Acad Sci U S A 102:7988-7993.

66. Owusu M, Annan A, Corman VM, Larbi R, Anti P, Drexler JF, Agbenyega O, Adu-Sarkodie Y, Drosten C. 2014. Human coronaviruses associated with upper respiratory tract infections in three rural areas of Ghana. PLoS One 9:e99782.

67. Mickleburgh S, Waylen K, Racey P. 2009. Bats as bushmeat: a global review. Oryx 43:217-234.

68. Anti P, Owusu M, Agbenyega O, Annan A, Badu EK, Nkrumah EE, Tschapka M, Oppong S, Sarkodie YA, Drosten C. 2015. Human–Bat interactions in rural West Africa. Emerg Infect Dis 21.

69. da Silva Filho LV, Zerbinati RM, Tateno AF, Boas LV, de Almeida MB, Levi JE, Drexler JF, Drosten C, Pannuti CS. 2012. The differential clinical impact of human coronavirus species in children with cystic fibrosis. J Infect Dis 206:384-388.

70. Chibo D, Birch C. 2006. Analysis of human coronavirus 229E spike and nucleoprotein genes demonstrates genetic drift between chronologically distinct strains. J Gen Virol 87:1203-1208.

APPENDIX 2
Deception down under: is Australia a hot spot for deception?

Image: Florian Gloza-Rausch

M. E. Herberstein, H. J. Baldwin and A. C. Gaskett. (2013) *Behavioral Ecology*, 25(1):12-16

25 YEARS
Behavioral Ecology
The official journal of the
ISBE
International Society for Behavioral Ecology

Behavioral Ecology (2014), 25(1), 12–16. doi:10.1093/beheco/art105

Invited Ideas

Deception down under: is Australia a hot spot for deception?

Marie E. Herberstein,[a] Heather J. Baldwin,[a] and Anne C. Gaskett[b]
[a]Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia and [b]School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

Received 21 November 2012; revised 7 October 2013; accepted 14 October 2013.

The Australian continent is renowned for its idiosyncratic flora and fauna and high diversity of endemic taxa (e.g., *Eucalyptus*, marsupials, and monotremes; Braithwaite RW. 1990. Australia's unique biota: implications for ecological processes. J Biogeogr. 17:347–354.). Given this diversity, it is perhaps not surprising that Australia is a coveted and productive field site for behavioral ecologists worldwide. The prevalence of some unusual animal behaviors is well documented, such as cooperative breeding in birds, low rates of herbivory, and high rates of pollination by vertebrates. However, other behavioral phenomena, especially those involving deception and exploitation, are also remarkably prevalent in some systems and still require comprehensive treatment. We examine 3 distinct forms of deception in entirely different taxa, cuckoos, crab spiders, and orchids, where there is strong evidence that deception is more prevalent in Australia than in other geographic regions. We offer several explanations addressing environmental conditions, evolutionary isolation, the prevalence of behavioral ecologists, and the research culture in Australia. The aim of this "idea" paper is to draw attention to intriguing patterns of deception in a limited number of well-studied systems and to generate several testable predictions. It is not intended as a thorough review of all deceptive systems, but we hope to stimulate more research, a systematic review, and further testing in this area.

Key words: **crab spider, cuckoo, deception, evolution, orchid.**

BROOD-PARASITIC CUCKOOS

Despite being substantially smaller than Africa (one-quarter) or Asia (one-sixth), Australia is home to ~20% of the world's brood-parasitic cuckoo species (Australia: 12, Africa: 13, Asia ~26 of 52 species in total; Davies 2010). These cheats outsource parental care and have evolved many deceptive signaling strategies to impair host detection including mimicry of host egg and chick colors and calls (Stoddard and Stevens 2010; Kilner and Langmore 2011; Langmore et al. 2011). Not only are brood-deceptive cuckoos more common than expected in Australia, the cuckoo–host relationship seems to differ from other regions. Australian cuckoo–host interactions are characterized by lower rates of cuckoo egg rejection by hosts in comparison to European counterparts (Langmore et al. 2003). Among other factors, Australia's milder environmental conditions provide hosts with longer breeding seasons allowing multiple clutches, whereas European hosts that are limited to single clutches may be under stronger selection for detection and rejection (Langmore et al. 2005).

Address correspondence to M.E. Herberstein. E-mail: marie.herberstein@mq.edu.au.
H.J. Baldwin is now at the Institute of Experimental Ecology, University of Ulm, 89069 Ulm, Germany.

DECEPTION BY CRAB SPIDERS

Crab spiders typically sit on flowers and hunt pollinators. European and American species are renowned for their cryptic coloration against the flower background (Théry and Casas 2002; Morse 2007). Their foraging success depends on being undetectable to the approaching pollinator. Recent work on Australian crab spiders, however, has revealed that they are far from cryptic, reflecting high amounts of UV light, thereby forming a strong color contrast against the floral background (Heiling et al. 2003). The UV component of their body color acts as a deceptive signal that lures both native and introduced pollinators closer to the flower (Heiling and Herberstein 2004; Heiling et al. 2005; Llandres et al. 2011). A multispecies comparison has revealed that deceptive signaling via UV reflection is seemingly overrepresented in Australian species (at least 5 species from 4 different genera). In the intensively studied European species, UV reflection is absent (Herberstein et al. 2009). Similarly, no cases of UV reflection are known from American or African species, and only one report of deceptive signaling via UV reflection comes from India (Bhaskara et al. 2008).

SEXUAL DECEPTION BY ORCHIDS

Sexually deceptive orchids are pollinated by male insects fooled into sexual behavior with orchid flowers (Vereecken 2009; Gaskett

2011; Schiestl 2011). These orchids do not offer a nectar reward and attract pollinators by mimicking the scents, shapes, and colors of female insects (Ayasse 2006; Gaskett and Herberstein 2010; Peakall et al. 2010; Ayasse et al. 2011; Gaskett 2012). Deceptive signaling involves exploitation of their pollinator's innate sensory biases (Stejskal et al. 2009; Gaskett 2013). Within Europe, sexual deception is largely restricted to only 2 genera: almost all ~255 species of the genus *Ophrys* and 1 species of *Orchis* (Paulus 2006; Ayasse et al. 2011; Gaskett 2011, but see Vereecken et al. 2012). In Australia, sexual deception has not just radiated from 1 genus, it has evolved independently at least 6 times, resulting in at least 11 genera and several hundred sexually deceptive species. This represents almost 50% of known sexually deceptive orchid species (Kores et al. 2001; Gaskett 2011).

Based on these 3 unrelated systems, we show that deceptive strategies are well represented in Australia, perhaps more so than expected compared with other regions. In the following sections, we propose several mechanisms that might explain this overrepresentation, including overall species diversity, the prevailing ecological conditions, the isolation and invasion history of Australia, and the prevalent research culture in Australia.

IS THE PREVALENCE OF SOME DECEPTIVE SYSTEMS A REFLECTION OF SPECIES DIVERSITY?

It may be that the diversification of deceptive strategies is a function of overall species diversity. In this case, we expect the overall species diversity of birds, spiders, and orchids to be proportional to the diversity of cuckoos, sexually deceptive orchids, and deceptive crab spiders. Based on a 2009 estimate, Australia holds about 8–9% of the world's bird diversity (9990 described species worldwide of which 828 occur in Australia; Chapman 2009). The diversity of cuckoos in Australia however is double that of the overall bird species diversity. Similarly, of more than 40 000 described spider species, ~8% occur in Australia (Chapman 2009), whereas almost all examples (5/6 species) of deceptive crab spiders identified so far come from Australia. Finally, the overrepresentation of sexual deception in Australian orchids is clearly not the result of an overall greater diversity or abundance of orchids in general: Australia is home to only ~5% of the world's orchid diversity (1200 species out of a global total of ~25 000; Dressler 2005; Hopper 2009; Table 1). Overall, species diversity in Australian birds, orchids, or spiders does not offer a strong explanation as to why deception is prevalent in these systems.

DOES DECEPTION EVOLVE READILY IN AUSTRALIA?

Australian environmental conditions

Australia's dryness, poor soils, and frequent fires explain several biological phenomena including production of nutrient poor biomass (resulting in low rates of herbivory) and abundant nectar and sap (leading to pollination by larger than average animals; Orians and Milewski 2007). It is conceivable that these overriding environmental constraints have led to selection for behavioral strategies that minimize the costs of acquiring or producing resources, such as outsourcing of parental care, and floral rewardlessness (Dafni and Bernhardt 1990; Gardener and Gillman 2001; Orians and Milewski 2007). If this were the case, we would predict that deceptive systems are more common in habitats with harsher climates (dry and poor soil, high frequency of fire), both within Australia and in other regions. However, we have limited information to evaluate this prediction.

The interaction between climate, Australia's long history of fire, and biodiversity may be factors explaining the incidence of deception especially in orchids. For example, fragmentation of orchid habitat caused by fire regimes may select for gene flow across larger distances, which may be facilitated by the greater outcrossing associated with sexual deception. Outcrossing is also associated with faster seed germination—this may also improve survival through the harsh summer by promoting earlier establishment (Peakall and Beattie 1996; Johnson and Nilsson 1999). Nevertheless, it may be phylogenetic constraints rather than resource limitation that select for the unusual parental strategies of Australian birds, including cooperative breeding (Cockburn 1996). Similarly, deception and rewardlessness in orchids provides outcrossing benefits and is no longer considered an adaptation to resource limitation per se (Jersáková and Johnson 2006). To fully examine the effect of environmental conditions requires investigating the distribution and abundance of deceptive and nondeceptive strategies of orchids and brood parasites across different Australian climate zones as well as similar environmental conditions outside Australia.

Isolation and invasions

Australia's long history of evolutionary isolation and recent invasions may explain the radiation of deception more broadly. If this is the case, we predict that deceptive species are relatively recent arrivals in Australia that exploit naive endemic receivers. Australia separated from the Gondwanan land mass 35–40 million years ago (Weimboy 1988; McLoughlin 2001). This isolation and the favorable Mediterranean and tropical climates have resulted in high levels of biodiversity and endemism (Cowling et al. 1996; Peel et al. 2007). This underlying diversity could provide a wide variety of potential targets for deception. In addition, isolated or island populations are typically more vulnerable to exploitation by invading species (Reaser et al. 2007). The mixing of existing and invading biota can lead to the evolution of new symbiotic relationships (Mooney and Cleland 2001) and this may well include deception.

Following the more recent continental contact between Australia and Asia, there has been frequent invasion by tropical species from Asia (O'Hara and Poore 2000; Murphy et al. 2006). Invading species with the capacity for deception may rapidly disperse and speciate throughout the continent, taking advantage of the relatively naive local species. The success of cuckoos, crab spiders, and orchids may have been the result of such invasions. Australian cuckoo genera are monophyletic and shared with Asia (Sorenson and Payne 2002), tentatively supporting the idea of recent invasions. Australian orchids stem from the independent arrival of 2 closely related subtribes from the Southern Hemisphere tribe Diurideae (Kores et al. 2001). It is unclear whether these orchid subtribes originated in Australia or close to Asia. The biogeography of Australian spiders is unresolved but likely to be complex, consisting of both endemics and recent invaders from southeast Asia (Murphy et al. 2006). To thoroughly test this idea, however, we require biogeographic data on the evolutionary relationships of Australian deceptive taxa.

DOES AUSTRALIA'S INTELLECTUAL AND RESEARCH CULTURE ENCOURAGE DISCOVERY OF DECEPTION?

Interpreting any data on the frequency of an observed phenomenon requires knowledge of potential biases in reporting. It may

Table 1

Orchid genera with sexual deception, food deception, or providing food rewards in 2 key regions of diversity: Australia and the Mediterranean

	Sexually deceptive pollination	Food deceptive pollination	Rewarding pollination
Australia (~1200 orchid spp.[a])	~11 genera[b,c] *Arthrochilus, Caladenia, Calonema, Calochilus, Chiloglottis, Cryptostylis, Drakaea, Leporella, Pterostylis, Pterygodium, Spiculaea*	~9 genera[b,c] *Caladenia, Diuris, Elythranthera, Glossodia, Leptoceras, Pheladenia, Rhizanthella, Spathoglottis, Thelymitra*	~24 genera[b,c] *Acianthus, Acrotys, Bulbophyllum, Caladenia, Calandra, Cymbidium, Cyrtostylis, Dendrobium, Dipodium, Diuris, Erythrorchis, Eulophia, Gastrodia, Genoplesium, Goodyera, Habenaria, Lipoxis, Microtis, Petalochilus, Pterostylis, Pheladenia, Prasophyllum, Spiranthes, Thelymitra*
Mediterranean [? orchid spp.]	5 genera[b,d] (total for Europe = 3) *Ophrys, Orchis galilaea, Serapias lingua*	4 genera[b,c] (total for Europe = 7) *Anacamptis, Ophioglossum, Cypripedium, Dactylorhiza, Neotinea, Orchis, Serapias*	~15 genera[b] (total for Europe = 16) *Anacamptis, Chamorchis, Coeloglossum, Epipactis, Epipogium, Goodyera, Gymnadenia, Herminium, Himantoglossum, Limodorum, Neotia, Orchis, Platanthera, Spiranthes*
Reported worldwide (~25 000 spp.[b])	24 genera[b,e]	47 genera[f]	~10 666 species (= two-thirds of all orchids[b,g])

Species with obligate self-pollination, unknown, or other forms of deception are omitted.
[a]Hopper (2009).
[b]Gaskett (2011).
[c]This may be up to 25 genera depending on current taxonomic reevaluations (Jones 2006; Hopper 2009).
[d]van der Cingel (2001).
[e]Jones (2006).
[f]Jones et al. (2010).
[g]Vereecken et al. (2012) and Dafforge (2005).
[h]Dressler (2005).
[i]Vereecken et al. (2010).
[j]Jersáková et al. (2006).

Table 2

Descriptives of the higher education environments in Australia, Europe, United Kingdom, United States of America, Canada, and Japan, indicating Australia's surprisingly large proportion of behavioral ecologists and consequent contribution to the journal *Behavioral Ecology* given the relatively few research institutes

Region	Number of research institutes[a]	Number of universities ranked in top 500[b]	Establishment of oldest university	Mean % of papers published in *Behavioral Ecology* 2010 and 2011[c]	Mean % of ISBE membership 2010 + 2011[c]
Europe	1614	162	1088, Università di Bologna	29	24.3
United Kingdom	160	38	1167, Oxford University	16.5	11.3
United States of America	2032	150	1636, Harvard University	24.3	39.2
Canada	145	25	1663, Université Laval	8.5	4.3
Japan	567	21	1630, University of Tokyo	2.5	5
Australia	50	19	1850, University of Sydney	8	8.5

[a]Data from www.arwu.cn and http://www.shanghairanking.com/.
[b]Data from *ISBE Newsletter* 2012;24(2).
[c]Data from International Society for Behavioral Ecology membership list November 2010 and November 2011.

be that the popularity of certain research areas, such as behavioral ecology in Australia, leads to a higher than expected reporting of charismatic phenomena including deceptive systems.

Behavioral ecology is thriving in Australia given the number of research institutions compared with other nations (Table 2). Over the past few years (2010 and 2011), 8% of all papers published in *Behavioral Ecology* were authored by researchers at Australian institutions (Simmons 2012). In contrast, 16.3% of papers were from the United Kingdom, which has 3 times as many institutions as Australia. The number of Australian-led papers published in other behavioral journals were similar over the same time period (*Animal Behaviour*: 7.7%; *Behavioral Ecology & Sociobiology*: 7.5%; Source: Web of Science, accessed 6 May 2013), suggesting that the observed publication data are not necessarily a reflection of the composition of editorial boards (*Behavioral Ecology*: 6 Australian-based editors out of a board of 28, *Animal Behaviour*: 2/32 editors, and *Behavioral Ecology & Sociobiology*: 6/41 editors).

The relatively recent research culture in Australia (Table 2), compared with Europe, United Kingdom, and the Americas, may also encourage natural history–based research (e.g., Shine 1994). Long-term data sets are not yet available for Australian study systems and this may stimulate researchers to think creatively and pursue non-traditional questions such as the lack of monogamy in passerines

(Double et al. 1997), slave making of unrelated juveniles for cooperative breeding in choughs (Heinsohn 1991), traumatic insemination and penis fencing in hermaphrodites (Michels and Newman 1998), and reversal of sex roles in nuptial feeding (Arnqvist et al. 2003). Discovering and characterizing deceptive systems requires similarly innovative approaches, and a peer environment that supports investigation into the obscure and the cryptic. However, we require more thorough analyses of the number of publications on deceptive and nondeceptive systems, and a measure of how often papers report on the same deceptive system, to fully evaluate the potential of publication bias.

In conclusion, we show that in 3 unrelated systems, deceptive species are more prevalent in Australia than in other world regions. We find no support that the diversity in deception is a reflection of overall species diversity. However, it is likely that the combination of climate and isolation favors the evolution of deception in Australia and that the academic environment favors studying these systems. Thus, we argue that Australia provides a "perfect storm" for deception.

FUNDING

We thank the Department of Biological Sciences, Macquarie University for financial support.

We thank our anonymous reviewers and the *Behavioral Ecology* editors for helpful feedback and suggestions.

Forum editor: Sue Healy

REFERENCES

Arnqvist G, Jones TM, Elgar MA. 2003. Insect behaviour: reversal of sex roles in nuptial feeding. Nature. 424:387.

Ayasse M. 2006. Floral scent and pollinator attraction in sexually deceptive orchids. In: Dudareva N, Pichersky E, editors. Biology of floral scent. Boca Raton (FL): CRC Press, Taylor & Francis Group. p. 219–241.

Ayasse M, Stökl J, Francke W. 2011. Chemical ecology and pollinator-driven speciation in sexually deceptive orchids. Phytochemistry. 72:1667–1677.

Bhaskara R, Brijesh CM, Ahmed S, Borges R. 2009. Perception of ultraviolet light by crab spiders and its role in selection of hunting sites. J Comp Physiol A. 195:409–417.

Bradshaw RHK. 1990. Australia's unique biota: implications for ecological processes. J Biogeogr. 17:347–354.

Chapman AD. 2009. Numbers of living species in Australia and the world. Australian Biological Resources Study. 2nd ed. Toowoomba (Australia): Department of the Environment, Water, Heritage and the Arts, Australian Government. ISBN [online] 978 0 642 56860 8.

van der Cingel NA. 2001. An atlas of orchid pollination: America, Africa, Asia and Australia. Rotterdam (The Netherlands): A.A. Balkema Publishers.

Cockburn A. 1996. Why do so many Australian birds cooperate: social evolution in the Corvida? In: Floyd R, Sheppard A, de Barro P, editors. Frontiers of population ecology. Melbourne (Australia): CSIRO Publishing. p. 451–472.

Cowling RM, Rundel PW, Lamont BB, Kalin Arroyo M, Arianoutsou M. 1996. Plant diversity in Mediterranean-climate regions. Trends Ecol Evol. 11:362–366.

Dafni A, Bernhardt P. 1990. Pollination of terrestrial orchids of southern Australia and the Mediterranean region. Evol Biol. 24:193–252.

Davies NB. 2010. Cuckoos, cowbirds and other cheats. London: T & AD Poyser.

Delforge P. 2005. Orchids of Europe, North Africa and the Middle East. London: A & C Black Publishers Ltd.

Double MC, Dawson D, Burke T, Cockburn A. 1997. Finding the fathers in the least faithful bird: a microsatellite-based genotyping system for the superb fairy-wren *Malurus cyaneus*. Mol Ecol. 6:691–693.

Dressler RL. 2005. How many orchid species? Selbyana. 26:155–158.

Gardener MC, Gillman MP. 2001. The effects of soil fertilizer on amino acids in the floral nectar of corncockle, *Agrostemma githago* (Caryophyllaceae). Oikos. 92:101–106.

Gaskett AC. 2011. Orchid pollination by sexual deception: pollinator perspectives. Biol Rev. 86:33–75.

Gaskett AC. 2012. Floral shape mimicry and variation in sexually deceptive orchids with a shared pollinator. Biol J Linn Soc. 106:469–481.

Gaskett AC. 2013. Colour and sexual deception in orchids: progress towards understanding the functions and pollinator perception of floral colour. In: Bernhardt P, Meyer R, editors. Darwin's orchids then and now. Chicago (IL): University of Chicago Press.

Gaskett AC, Herberstein ME. 2010. Colour mimicry and sexual deception by Tongue orchids (*Cryptostylis*). Naturwissenschaften. 97:97–102.

Heiling AM, Cheng K, Chittka L, Goeth A, Herberstein ME. 2005. The role of UV in crab spider signals: effects on perception by prey and predators. J Exp Biol. 208:3925–3931.

Heiling AM, Herberstein ME. 2004. Predator-prey coevolution: Australian native bees avoid their spider predators. Proc R Soc B. 271(Suppl 4):S196–S198.

Heiling AM, Herberstein ME, Chittka L. 2003. Crab-spiders manipulate flower signals. Nature. 421:334.

Heinsohn RG. 1991. Kidnapping and reciprocity in cooperatively breeding white-winged choughs. Anim Behav. 41:1097–1100.

Herberstein ME, Heiling AM, Cheng K. 2009. Evidence for UV-based sensory exploitation in Australian but not European crab spiders. Evol Ecol. 23:621–634.

Hopper SD. 2009. Taxonomic turmoil down-under: recent developments in Australian orchid systematics. Ann Bot. 104:447–455.

Jersáková J, Johnson S. 2006. Lack of floral nectar reduces self-pollination in a fly-pollinated orchid. Oecologia. 147:60–68.

Jersáková J, Johnson SD, Kindlmann P. 2006. Mechanisms and evolution of deceptive pollination in orchids. Biol Rev. 81:219–235.

Johnson SD, Nilsson LA. 1999. Pollen carryover, geitonogamy, and the evolution of deceptive pollination systems in orchids. Ecology. 80:2607–2619.

Jones DL. 2006. A complete guide to native orchids of Australia including the Island Territories. French's Forest (Australia): Reed New Holland.

Jones DL, Hopley T, Duffy SM. 2010. Australian tropical rainforest orchids. Canberra (Australia): CSIRO.

Kilner RM, Langmore NE. 2011. Cuckoos versus hosts in insects and birds: adaptations, counter-adaptations and outcomes. Biol Rev. 86:836–852.

Kores PJ, Molvray M, Weston PH, Hopper SD, Brown AP, Cameron KM, Chase MW. 2001. A phylogenetic analysis of Diurideae (Orchidaceae) based on plastid DNA sequence data. Am J Bot. 88:1903–1914.

Langmore NE, Kilner RM, Butchart SHM, Maurer G, Davies NB, Cockburn A, Macgregor NA, Peters A, Magrath MJL, Dowling DK. 2005. The evolution of egg rejection by cuckoo hosts in Australia and Europe. Behav Ecol. 16:686–692.

Langmore NE, Stevens M, Maurer G, Heinsohn R, Hall ML, Peters A, Kilner RM. 2011. Visual mimicry of host nestlings by cuckoos. Proc R Soc B. 278:2455–2463.

Llandres AL, Gawryszewski FM, Heiling AM, Herberstein ME. 2011. The effect of colour variation in predators on the behaviour of pollinators: Australian crab spiders and native bees. Ecol Entomol. 36:72–81.

McLoughlin S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Aust J Bot. 49:271–300.

Michels NK, Newman LJ. 1998. Sex and violence in hermaphrodites. Nature. 391:647.

Mooney HA, Cleland EE. 2001. The evolutionary impact of invasive species. Proc Natl Acad Sci USA. 98:5446–5451.

Morse DH. 2007. Predator upon a flower. Cambridge (MA): Harvard University Press.

Murphy N, Framenau V, Donnellan S, Harvey M, Park Y, Austin A. 2006. Phylogenetic reconstruction of the wolf spiders (Araneae: Lycosidae) using sequences from the 12S rRNA, 28S rRNA, and NADH1 genes: implications for classification, biogeography, and the evolution of web building behavior. Mol Phylogenet Evol. 38:503–602.

O'Hara TD, Poore GCB. 2000. Patterns of distribution of southern Australian marine echinoderms and decapods. J Biogeogr. 27:1321–1335.

Orians GH, Milewski AV. 2007. Ecology of Australia: the effects of nutrient-poor soils and intense fires. Biol Rev. 82:393–423.

Paulus HF. 2006. Deceived males—pollination biology of the Mediterranean orchid genus *Ophrys* (Orchidaceae). J Eur Orch. 38:303–353.

Peakall R, Beattie AJ. 1996. Ecological and genetic consequences of pollination by sexual deception in the orchid Caladenia tentaculata. Evolution. 50:2207–2220.
Peakall R, Ebert D, Poldy J, Barrow RA, Francke W, Bower CC, Schiestl FP. 2010. Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually deceptive Chiloglottis orchids: implications for pollinator-driven speciation. New Phytol. 188:437–450.
Peel MC, Finlayson BL, McMahon TA. 2007. Updated world map of the Koppen-Geiger climate classification. Hydrol Earth Syst Sci. 11:1633–1644.
Reaser JK, Meyerson LA, Cronk Q, De Poorter M, Eldredge LG, Green E, Kairo M, Latasi P, Mack RN, Mauremootoo J, et al. 2007. Ecological and socioeconomic impacts of invasive alien species in island ecosystems. Environ Conserv. 34:98–111.
Schiestl FP. 2011. Pollination: sexual mimicry abounds. Curr Biol. 20:R1020–R1022.
Shine R. 1994. National peculiarities, scientific traditions and research directions. Trends Ecol Evol. 9:309.
Simmons LW. 2012. Behavioral Ecology—report from the editor-in-chief. ISBE Newslett. 24:2–4.
Soreinson MD, Payne RB. 2002. Molecular genetic perspectives on avian brood parasitism. Integr Comp Biol. 42:388–400.
Stoddard MC, Stevens M. 2010. Pattern mimicry of host eggs by the common cuckoo, as seen through a bird's eye. Proc R Soc B. 277:1387–1393.
Streinzer M, Paulus HF, Spaethe J. 2009. Floral colour signal increases short-range detectability of a sexually deceptive orchid to its bee pollinator. J Exp Biol. 212:1365–1370.
Théry M, Casas J. 2002. Predator and prey views of spider camouflage. Nature. 415:133.
Vereecken N. 2009. Deceptive behaviour in plants. I. Pollination by sexual deception in orchids: a host-parasite perspective. In: Baluska F, editor. Plant-environment interactions—from sensory plant biology to active behavior. Berlin (Germany): Springer-Verlag. p. 203–222.
Vereecken NJ, Dafni A, Cozzolino S. 2010. Pollination syndromes in Mediterranean orchids—implications for speciation, taxonomy and conservation. Bot Rev. 76:220–240.
Vereecken NJ, Wilson CA, Hötling S, Schulz S, Banketov SA, Mardulyn P. 2012. Pre-adaptations and the evolution of pollination by sexual deception: Cope's rule of specialization revisited. Proc R Soc B. 279:4786–4794.
Wemsby M. 1988. Comparing Australian ecosystems to those elsewhere. BioScience. 38:549–556.

MACQUARIE UNIVERSITY

ANIMAL RESEARCH AUTHORITY (ARA)

AEC Reference No.: 2011/061 **Date of Expiry:** 31 December 2012

Full Approval Duration: *1 January 2012 to 31 December 2014 (36 months)*

This ARA remains in force until the Date of Expiry (unless suspended, cancelled or surrendered) **and will only be renewed upon receipt of a satisfactory Progress Report before expiry / is contingent upon receipt of a Final Report at the end of this period (see Approval email for submission details).**

Principal Investigator:
Dr Adam Stow
Dept of Biological Sciences
0452 180 342
Adam.stow@mq.edu.au

Associate Investigators:
Heather Baldwin 0452 180 342

In case of emergency, please contact:
Australian Wildlife Conservancy, Newhaven Sanctuary Manager, Northern Territory, +61 8 9864 6000
or the Principal Investigator / Associate Investigators named above

The above-named are authorised by MACQUARIE UNIVERSITY ANIMAL ETHICS COMMITTEE to conduct the following research:

Title of the project: Connectivity, phylogeography and virology in African leaf-nosed bats

Type of animal research: 5: Research; Human or animal health and welfare
Aims of project:

1. Utilize genetic tools to characterize populations structure of the *Hipposideros caffer/ruber* complex in Ghana
2. Explore the relationship between host dynamics and coronavirus prevalence, diversity and evolution
3. Investigate the cost of coronavirus infection through an acoustic study

All procedures must be performed as per the AEC approved protocol, unless stated otherwise by the AEC and/or AWO.

Maximum numbers approved:

Species	Sex	Procedure	*Year 1*	*Year 2*	*Year 3*	Total	Supplier/ Source
Hipposideros caffer/ruber	M/F	*Capture, measurements, wing biopsy, faecal sample , acoustic recording (400 in each of 8 colonies, twice yearly)*	6400	6400	6400	**19200**	Wild
					TOTAL	**19200**	

Location of research:
Ghana, West Africa: Kwamang village, Ahanti Region; Buoyem village, Brong-Akpafu region; Forikrom, Brong-Akpafu region; Likpe-Todome, Volta region; Okplo mine, Akpafu-Todzi, Volta region; Elmina Castle, Elmina, Central region

Amendments since initial approval: N/A

Conditions of Approval: N/A

*All Permits/Licenses (to obtain and use fauna; to conduct research at interstate/overseas locations; to house animals, etc.) must be obtained **prior** to work commencing, and copies forwarded to the Animal Ethics Secretariat.*

Being animal research carried out in accordance with the Code of Practice for a recognised research purpose and in connection with animals (other than exempt animals) that have been obtained from the holder of an animal supplier's license.
This authority remains in force from ***1 January 2012 to 31 December 2012***, unless suspended, cancelled or surrendered, **and will only be renewed upon receipt of a PROGRESS REPORT annually.**

Prof Michael Gillings (Chair, Animal Ethics Committee)

Approval Date: ***8 December 2011***

Adapted from Form C (issued under part IV of the Animal Research Act, 1985)

EXECUTIVE DIRECTOR
WILDLIFE DIVISION
FORESTRY COMMISSION
ACCRA

Receipt No A0499 of 28-04-2009

FCWD/GH-01

APPLICATION TO STUDY OR CONDUCT RRSEARCH WITHIN WILDLIFE PROTECTED AREAS IN GHANA

APPLICANT'S INFORMATION

First name: Samuel
Last (family) name: OPPONG
Nationality of passport: GHANAIAN
Passport Number: H 1426280
Date of birth: 29/11/57
Place of birth: Kumasi
Highest degree earned: Ph D (Professor of Wildlife & Range Mgt)
Specialized field of research: Wildlife and Range Mgt
Affiliated institution: FACULTY OF RENEWABLE NATURAL RESOURCES
Institution address: FACULTY OF RENEWABLE NATURAL RESOURCES KNUST, KUMASI
Current address: FRNR, KNUST, KUMASI
Email: kobbyoppong@yahoo.com
Telephone: 051 60381
Fax: 051 60375

Information on proposed project

Title: VIRUSES IN GHANAIAN BATS
Proposed Start Date: 22/01/09 End Date: 21/01/12

Specimen and /or samples [ ] will / [ ] will not be collected within wildlife protected areas in Ghana during the course of the proposed project

Name and region of proposed research sites in protected areas of Ghana

1. Bonyere Caves
2. Kintosi Zoological gardens
3. Kwadwoma Bat Caves
4. Bosomtwe Area - Ashanti Region
5. Shai Hills - Greater Accra Region

To: Executive Director, Wildlife Division, Forestry Commission, P.O. Box MB 239 Accra

If I and my collaborators receive permission to carry out the proposed project, we will comply with any and all laws, regulations, stipulations, and agreements germane to this work, and, in addition, abide by any additional orders and requests made by government officials.

I have submitted the following documents with this application for your consideration:

1. project proposal
2. biographical sketches for each of the 1 (number of persons) collaborators
3. A photocopy of the signature page from the passport (1)

Sincerely yours,

22/01/09
Signature date

I.S.K. Ofosu-Gyei
Printed

Project Leader

EXECUTIVE DIRECTOR
WILDLIFE DIVISION
FORESTRY COMMISSION
ACCRA
26/1/09

Official use only
Research project code
03/01/2009

This research/ study is approved under the following conditions:

1. Pay a flat research fee of $300 or its equivalent in cedis
2. Pay for entrance fee, guide fees and other services that you may require
3. Submit one (1) copy of thesis/report/document to the Executive Director.
4. Provide three (3) copies of any publication emanating from the study.
5. Obtain CITES export permit should there be the need to send specimen outside the country

FOR: EXECUTIVE DIRECTOR
NANA AFUA ABU-NYAH

EXECUTIVE DIRECTOR
WILDLIFE DIVISION
FORESTRY COMMISSION
ACCRA