A peer-reviewed version of this preprint was published in PeerJ on 5 November 2019.

View the peer-reviewed version (peerj.com/articles/8013), which is the preferred citable publication unless you specifically need to cite this preprint.

Näpflin K, O'Connor EA, Becks L, Bensch S, Ellis VA, Hafer-Hahmann N, Harding KC, Lindén SK, Olsen MT, Roved J, Sackton TB, Shultz AJ, Venkatakrishnan V, Videvall E, Westerdahl H, Winternitz JC, Edwards SV. 2019. Genomics of host-pathogen interactions: challenges and opportunities across ecological and spatiotemporal scales. PeerJ 7:e8013 https://doi.org/10.7717/peerj.8013

Genomics of hosts-pathogen interactions: challenges and opportunities across ecological and spatiotemporal scales

Kathrin Näpflin [Corresp., 1], **Lutz Becks** [2], **Staffan Bensch** [3], **Vincenzo A Ellis** [3], **Nina Hafer-Hahmann** [4, 5], **Karin C Harding** [6, 7], **Sara K Lindén** [8], **Emily A O'Connor** [3], **Morten T Olsen** [9], **Jacob Roved** [3], **Timothy B Sackton** [10], **Allison J Shultz** [11], **Vignesh Venkatakrishnan** [8], **Elin Videvall** [3, 12], **Helena Westerdahl** [3], **Jamie C Winternitz** [4, 13], **Scott V Edwards** [1, 7]

[1] Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, United States of America

[2] Limnological Institute, Universität Konstanz, Konstanz, Germany

[3] Department of Biology, Molecular Ecology and Evolution Lab, Lund University, Lund, Sweden

[4] Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Plön, Germany

[5] EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

[6] Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden

[7] Gothenburg Centre for Advanced Studies in Science and Technology, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

[8] Department of Medical Chemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[9] Section for Evolutionary Genomics, Natural History Museum of Denmark, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[10] Informatics Group, Harvard University, Cambridge, Massachusetts, United States of America

[11] Ornithology Department, Natural History Museum of Los Angeles County, Los Angeles, California, United States of America

[12] Center for Conservation Genomics, Smithsonian Conservation Biology Institute, National Zoological Park, Washington, District of Columbia, United States of America

[13] Department of Animal Behaviour, Universität Bielefeld, Bielefeld, Germany

Corresponding Author: Kathrin Näpflin
Email address: knaepflin@fas.harvard.edu

Evolutionary genomics has recently entered a new era in the study of host-pathogen interactions. A variety of novel genomic techniques has transformed to the identification, detection and classification of both hosts and pathogens, allowing a greater resolution that helps decipher their underlying dynamics and provides novel insights into their environmental context. Nevertheless, many challenges to a general understanding of host-pathogen interactions remain, in particular in the synthesis and integration of concepts and findings across a variety of systems and different spatiotemporal and ecological scales. In this perspective we aim to highlight some of the commonalities and complexities across diverse studies of host-pathogen interactions, with a focus on ecological, spatiotemporal variation, and the choice of genomic methods used. We performed a quantitative review of recent literature to investigate links, patterns and potential tradeoffs between the complexity of genomic, ecological and spatiotemporal scales undertaken in individual host-pathogen studies. We found that the majority of studies used whole genome resolution to address their research objectives across a broad range of ecological

scales, especially when focusing on the pathogen side of the interaction. Nevertheless, genomic studies conducted in a complex spatiotemporal context are currently rare in the literature. Because processes of host-pathogen interactions can be understood at multiple scales, from molecular-, cellular-, and physiological-scales to the levels of populations and ecosystems, we conclude that a major obstacle for synthesis across diverse host-pathogen systems is that data are collected on widely diverging scales with different degrees of resolution. This disparity not only hampers effective infrastructural organization of the data but also data granularity and accessibility. Comprehensive metadata deposited in association with genomic data in easily accessible databases will allow greater inference across systems in the future, especially when combined with open data standards and practices. The standardization and comparability of such data will facilitate early detection of emerging infectious diseases as well as studies of the impact of anthropogenic stressors, such as climate change, on disease dynamics in humans and wildlife.

Genomics of hosts-pathogen interactions: challenges and opportunities across ecological and spatiotemporal scales

Authors and Affiliations

1 Näpflin, Kathrin
2 Becks, Lutz
3 Bensch, Staffan
3 Ellis, Vincenzo A.
4, 5 Hafer-Hahmann, Nina
6, 7 Harding, Karin C.
8 Lindén, Sara K.
3 O'Connor, Emily A.
9 Olsen, Morten T.
3 Roved, Jacob
10 Sackton, Timothy B.
11 Shultz, Allison J.
8 Venkatakrishnan, Vignesh
3,12 Videvall, Elin
3 Westerdahl, Helena
4, 13 Winternitz, Jamie C.
1, 6 Edwards, Scott V.

1 Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA
2 Limnological Institute University Konstanz, Aquatic Ecology and Evolution, Konstanz, Germany
3 Department of Biology, Molecular Ecology and Evolution Lab, Lund University, Lund, Sweden
4 Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Plön, Germany
5 EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland
6 Gothenburg Centre for Advanced Studies in Science and Technology, Chalmers University of Technology and University of Gothenburg, Göteborg, Sweden
7 Department of Biological and Environmental Sciences, Gothenburg University, Sweden
8 Department of Medical Chemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
9 Section for Evolutionary Genomics, Natural History Museum of Denmark, Department of Biology, University of Copenhagen, Denmark
10 Informatics Group, Harvard University, Cambridge, MA, USA
11 Ornithology Department, Natural History Museum of Los Angeles County, Los Angeles, CA, USA.
12 Center for Conservation Genomics, Smithsonian Conservation Biology Institute, National Zoological Park, Washington, DC, USA.
13 Department of Animal Behaviour, Bielefeld University, Bielefeld, Germany

Abstract

Evolutionary genomics has recently entered a new era in the study of host-pathogen interactions. A variety of novel genomic techniques has transformed to the identification, detection, and classification of both hosts and pathogens, allowing a greater resolution that helps decipher their underlying dynamics and provides novel insights into their environmental context. Nevertheless, many challenges to a general understanding of host-pathogen interactions remain, in particular in the synthesis and integration of concepts and findings across a variety of systems and different spatiotemporal and ecological scales. In this perspective, we aim to highlight some of the commonalities and complexities across diverse studies of host-pathogen interactions, with a focus on ecological, spatiotemporal variation, and the choice of genomic methods used. We performed a quantitative review of recent literature to investigate links, patterns and potential tradeoffs between the complexity of genomic, ecological and spatiotemporal scales undertaken in individual host-pathogen studies. We found that the majority of studies used whole-genome resolution to address their research objectives across a broad range of ecological scales, especially when focusing on the pathogen side of the interaction. Nevertheless, genomic studies conducted in a complex spatiotemporal context are currently rare in the literature. Because processes of host-pathogen interactions can be understood at multiple scales, from molecular-, cellular-, and physiological-scales to the levels of populations and ecosystems, we conclude that a major obstacle for synthesis across diverse host-pathogen systems is that data are collected on widely diverging scales with different degrees of resolution. This disparity not only hampers effective infrastructural organization of the data but also data granularity and accessibility. Comprehensive metadata deposited in association with genomic data in easily accessible databases will allow greater inference across systems in the future, especially when combined with open data standards and practices. The standardization and comparability of such data will facilitate early detection of emerging infectious diseases as well as studies of the impact of anthropogenic stressors, such as climate change, on disease dynamics in humans and wildlife.

Subjects: Biodiversity, Ecology, Evolutionary Studies, Genomics, Zoology

Keywords: *Plasmodium*, MHC, immunotoxins, mucus, natural selection, GWAS, infectious diseases, anthropogenic stressors, co-evolution, epidemiological surveillance

Introduction

Pathogens are widely agreed to be among the strongest agents of natural selection in nature, and their influence on the genomes of host species is often readily detectable (Kosiol et al., 2008; Enard et al., 2016; Quach et al., 2016; Shultz & Sackton, 2019). With the advent of rapid DNA sequencing technologies, genetic studies of host-pathogen interactions have moved from single gene perspectives to genome-wide approaches interrogating whole genomes of hosts and/or pathogens. At the same time, these studies have begun to tackle an increasingly diverse array of systems in both the field and laboratory, and have expanded from analysis of single pathogens to multiple pathogens under a variety of conditions. Environmental factors and gene-by-environment interactions, such as those beginning to be studied in microbiome research (Libertucci & Young, 2019), are increasingly appreciated as important in modulating the severity and fitness consequences of infections (Sekirov et al., 2010; Kamada et al., 2013; Villarino et al., 2016; Chomwong et al., 2018). As genomic approaches become increasingly accessible and affordable, it is becoming clear that the limiting factor in host-pathogen research is often not the technical aspects of sequencing pathogens or host genomes, but rather the ecological, immunological and epigenetic context in which genomic data are embedded (Kratochwil & Mayer, 2015). To mention one example, post-translational modifications of proteins in the mucus are known to play critical roles in pathogen defense in addition to genetic factors (Linden et al., 2008).

Host-pathogen studies encompass an extraordinary variety of temporal and spatial scales, including wide ranges of ecological settings and pathogen complexities - such as experimental versus field studies or single versus multiple pathogens - as well as genomic complexities, ranging from candidate gene studies to whole genome scans (Fig. 1A). Any single study can be classified according to these scales, with concomitant benefits and deficiencies in capturing the details of host-pathogen interactions in the real world. For example, studies aiming to link the evolution of host and pathogen genomes and to detect genomic signatures of host-pathogen interactions have arisen from searches for associations with single host candidate genes, such as genes of the major histocompatibility complex (MHC) (Hill et al., 1991; Kaslow et al., 1996; Wegner, Reusch & Kalbe, 2003; Meyer-Lucht & Sommer, 2005; Savage & Zamudio, 2011), to genome-wide scans for associations with resistance or susceptibility (e.g., (Fumagalli et al., 2011; Bartha et al., 2013). We now have genomic insights into host-pathogen interactions that stem from field studies investigating temporal and spatial patterns (Hill et al., 1991; Savage & Zamudio, 2011; Penczykowski, Laine & Koskella, 2015; Bourgeois et al., 2017); to

experimentally evolving populations or ancient DNA studies encompassing hundreds or thousands of host generations (Bos et al., 2011; Cairns et al., 2017; Tso et al., 2018; Spyrou et al., 2019); to phylogenetic and comparative studies spanning tens of millions of years (Enard et al., 2016; Koonin, Makarova & Wolf, 2017; Shultz & Sackton, 2019). This variety makes it challenging to draw broad generalizations linking processes on different scales and to date, few syntheses have attempted to bridge the many temporal and spatial scales on which host-pathogen studies take place.

In this perspective, we aim to address the complexities and commonalities of diverse studies of host-pathogen interactions through the lens of evolutionary genomics. We emphasize the wide range of approaches used recently and focus primarily on evolutionary responses of hosts to pathogens (Fig. 1). We first document the diversity of recent studies of host-pathogen interactions through a comprehensive analysis of the recent literature on the subject. This survey documents the sheer diversity of temporal and spatial scales on which host-pathogen studies are conducted, but also reveals that the heterogeneity of results across studies, from laboratory to field to experimental settings, poses a challenge for synthesis. Our survey identifies gaps in emphasis on research on host-pathogen interactions, but also reveals opportunities for discovering common principles and methodologies that are likely to drive the research field forward. We then review major themes in the study of interactions between hosts and pathogens in the wild. While daunting in terms of confounding variables, such studies provide opportunities for studying the synergistic effects of anthropogenic change and the evolutionary response to epizootics. At the same time, an increasing number of experimental studies that examine the effects of multiple interacting pathogens on their hosts, or of host microbiome on infection outcome, capture some of the reality of epizootics in nature. We conclude that the full promise of genomic and other -omics approaches to further our understanding of host-pathogen interactions will not be realized until these data are thoroughly and consistently embedded in high quality, consistent, and reproducible ecological and environmental metadata. Increased resolution of ecological metadata, smart databases that facilitate collaboration and comparisons across studies, and deposition of voucher specimens associated with specific studies in museum collections are just some of the ways in which genomic data can realize their full potential. These new tools will facilitate the application of knowledge of basic principles of host-pathogen interactions to real world problems affecting wildlife, endangered species, and ultimately human welfare.

Survey methodology

This perspective piece is the result of discussions held during the "Origins of Biodiversity Workshop" organized during May 29 – June 2, 2017 by Chalmers University of Technology and the University of Gothenburg, Sweden, under the auspices of the Gothenburg Centre for Advanced Studies (GoCAS). We gathered international scholars and students with interdisciplinary backgrounds to discuss future perspectives of the study of host-pathogen co-evolution in the genomic era. During the workshop we identified major directions that have been enabled by advances in genomic techniques and in particular we highlight the resulting diversity of host-pathogen studies in their ecological, temporal and genomic detail at which they are studied. Our goal is not to provide a complete overview of the host-pathogen literature, but rather illustrate the diversity of research undertaken in the field and the associated challenges towards a comparable and inter-communicative understanding of causes and consequences of host-pathogen interactions across systems. To quantify currently studied dimensions (ecological complexity, spatiotemporal scope and genomic scale; Fig. 1A) of host-pathogen research we conducted a literature search on Web of Science (accessed August 30, 2018) with the following search query: (host-parasite* OR host-pathogen*) AND (genomic*). We refined the search hits by document type to include only articles, covering the publication years 2014-2018, and excluding studies with no genomic aspect. We scored 263 papers based on broadly defined categories for each scale defined in Table 1. The reference list and scoring results can be found in Supplementary Table 1. We used Spearman's rho to assess the rank based association between scales and adjusted p-values for multiple testing (Benjamini & Hochberg, 1995).

Understanding the diversity of host-pathogen studies across genomic, ecological and spatiotemporal scales

We have outlined in the introduction that the published literature on host-pathogen interaction spans a diverse range of genomic, ecological and spatiotemporal scales. However, how the current published literature is distributed within this multidimensional space has not been mapped out (Fig. 1A). We thus first performed a literature search to classify and quantify the distribution of studies across these three scales. To understand the range of investigation at the genomic, ecological and spatiotemporal scale in recent studies on host-pathogen systems, and to discern where gaps in recent efforts might persist, we reviewed 263 studies of host-pathogen interactions published in the period between 2014-2018 (see Table 1 for scoring categories and Survey Methodology for specific details). A better understanding of the current placement of

host-pathogen studies should help us gain a better insight on how genomics can contribute to the understanding of host-pathogen interactions that are being studied from the perspectives of hosts or pathogens but also at such various level of biological detail.

We scored each study on three scales: genomic complexity, temporal and spatial complexity, and ecological complexity (Table 1). We found that high genomic resolution (mean score = 5.4 ± 2.1 SD, range = 1-7) at the level of the whole genome is employed to investigate questions that span the whole range of ecological scales, from theoretical to multi-species natural systems with environmental variability (mean score = 4.1 ± 2.4 SD, range = 1-9; Fig. 1B). Investigations of pathogen genomics dominate the dataset, whereas genomic investigations of hosts are less common and more often examine a reduced representation of the genome, transcriptome, or proteome. Studies encompassing the interaction of both host and pathogens simultaneously are rare. Genomic techniques are rarely used to address complex spatiotemporal scales (mean score = 4.1 ± 2.3 SD, range = 2-11), such as throughout the geographical range of a species, or across multiple different species. (Fig. 1C). Intriguingly, with increasing ecological complexity in a study, more complex spatiotemporal scales are also addressed (Fig. 1D). However, only a few studies are classified as complex in terms of spatiotemporal setting: Across all studies, spatial (mean score = 1.6 ± 1.0 SD, range = 1-5) and temporal scores (mean score = 2.4 ± 1.7 SD, range = 1-6) are on average low. In particular, studies of complex spatial scales, such as interrogation across multiple populations across a species' range, are virtually missing.

The evaluation of published studies on host-pathogen systems not only reveals the expected recent increase in whole genome datasets for a broad range of host-pathogen studies, but also the gaps in addressing complex systems on ecological and spatiotemporal scales. Perhaps most critically, we suspect that the quantity and complexity of the sequence data in many recent and ongoing investigations of host-pathogen interactions pose increasing challenges for comparisons across studies. The lack of comprehensive cross-taxon comparative databases of host-pathogen interactions likely impedes the synthesis of individual host-pathogen studies and translation of new knowledge into solutions for real world problems. In the real-world, (a) pathogens attack hosts in the context of changing host environments, (b) these environments are increasingly impacted by anthropogenic forces such as climate change, and (c) are usually characterized by diverse communities of pathogenic and non-pathogenic

organism. Our cross-section of recent studies of host-pathogen interactions suggests that these complexities are rarely captured in a single study.

Disentangling hidden histories in genes and genomes of hosts and pathogens

The pathogenic lifestyle is ubiquitous across the tree of life, and pathogens are estimated to represent a substantial proportion of the diversity and biomass of many ecosystems (Padra et al., 2018). The advent of high throughput sequencing has facilitated the discovery of numerous previously unknown pathogens and complex host-pathogen life histories, while metagenomic approaches have allowed for the classification of host- and pathogen-associated microbial communities, which have often been linked to successful pathogen infection and disease development (Sekirov et al., 2010; Kamada et al., 2013). This large increase in the number of observed pathogen strains and species along with associated microbes suggests that a pathogen rarely occurs alone, and instead may commonly be a member of a larger community (Robinson, Bohannan & Young, 2010; Schmid-Hempel, 2011; Gregory et al., 2019). Hence, understanding the interplay between multiple pathogens and associated microbiomes requires disentangling several levels of complexity. It is also is crucial to gain an understanding of the fitness effects of each putative pathogen on its host, because the magnitude of the fitness cost (i.e. virulence) of a pathogen during infection determines its place on the mutualist-pathogen continuum. In principal, this requires demonstration of a fitness cost to the host, yet demonstrating fitness effects of many putative pathogens in nature is challenging and often requires datasets that are much larger than those obtained in a typical field study [see Box 1]. Importantly, the fact that measuring the fitness consequences of infections in wild animals is challenging does not imply that pathogens are insignificant selective agents in the evolution of host genomes. In fact, pathogens are widely regarded to be among the strongest selective agents. Mutations conferring moderate or large benefits of resistance to hosts can become readily fixed by selection and are detectable through genome scans (Nielsen, 2005; Vitti, Grossman & Sabeti, 2013; Haasl & Payseur, 2016). Thus, comparative and population genetic studies of host genomes present compelling approaches for studying the presumed impact of pathogens (Fig. 2).

Genetic variation is typically studied at different levels, such as across species (Fig. 2A), across populations (Fig. 2B), within populations (Fig. 2C) or through time (Fig. 2D) to

disentangle the underlying genetics of host-pathogen interactions. For this purpose, two main approaches are typically employed. On the one hand, the underlying genetic architecture can be inferred using genotype-phenotype association studies. The statistical association between genomic loci and host-pathogen phenotypes is interpreted as evidence for the underlying genetics of a given phenotype (Hirschhorn et al., 2002). On the other hand, instead of determining fitness costs of pathogens in single experiments or surveys (see also Box 1), biologists have turned to signals of natural selection over evolutionary time as recorded in host genomes (Sabeti et al., 2006). While these genome scans typically cannot directly test the causal selective agent, they do provide insight into the possible biological processes that are adapting most rapidly in host genomes (Biswas & Akey, 2006). Indeed, analysis of signatures of selection in host genomes identified pathogens as the most likely drivers of the observed patterns in a number of studies. For example, in *Drosophila*, Sackton and co-authors (2007) identified that a class of immune genes that directly interact with pathogens, such as receptor genes, exhibited a high proportion of genes under positive selection compared to genome-wide observations. Similarly, across mammals, viral interacting proteins have stronger signals of adaptation than other protein-coding genes across the genome (Enard et al., 2016), and more of these genes than expected by chance are also evolving by positive selection in birds (Shultz & Sackton, 2019).

A combination of selection scans and association studies has revealed important insights into differences in infectious disease susceptibilities, the identification of specific protective genes and alleles, and their evolutionary origin in humans, the most intensely studied organism with respect to disease (Nielsen et al., 2005; Kwiatkowski, 2005; Williams et al., 2005; Karlsson, Kwiatkowski & Sabeti, 2014; Malaria Genomic Epidemiology Network et al., 2015; Enard et al., 2016; Enard & Petrov, 2018). There have been similar advances in the understanding of the underlying genetics of natural host-pathogen systems in the wild. For example, Bourgeois *et al*. (2017) was able to confirm and refine previously identified quantitative trait loci that confer resistance in the planktonic crustacean *Daphnia magna* to the pathogen *Pasteuria ramosa.* Furthermore, investigations of signals of selection have identified additional genomic regions consistent with the evolution of resistance that were not identified by association approaches. Such loci present further candidates moderating the host-pathogen interactions, but without a clear association with specific phenotypic traits, evolution in response to other environmental variables correlated with pathogens often cannot be excluded (Bourgeois et al., 2017).

Despite the success of genome-wide associations and selection scans to find genomic evidence of pathogen pressures on hosts, simultaneous genomic investigation of the co-evolutionary dynamics between host and pathogen within a single system remain rare (see Fig. 1). Indeed, today, few systems have the genomic resources available to be truly interrogate ongoing genomic changes between pathogen and host in parallel. One such example is Bartha *et al.* 2013, who identified linked sequence variants between humans and HIV through genome-wide-association scans. This study highlighted both host and viral loci that are potentially involved in the co-evolutionary dynamics between host and pathogen. Additionally, emerging studies of experimental evolution in the field or laboratory, or multigenerational sampling of natural populations of hosts and pathogens, have successfully identified novel adaptive alleles in both hosts (Scanlan et al., 2015; Cairns et al., 2017) and pathogens (Pal et al., 2007). The reciprocal nature of the interactions between host and pathogens over time naturally lead researchers to ask whether host and pathogen species co-diversify over evolutionary time and to what extent genomics can inform the underlying processes. Indeed, attempts to detect co-speciation among hosts and pathogens date back to the beginning of the 20th century (reviewed in (Vienne et al., 2013). However, inferring co-evolutionary history through comparisons of host and pathogen phylogenies is challenging. For example, such comparisons can mistake a host shift followed by co-diversification as co-speciation (Vienne, Giraud & Shykoff, 2007). The former mechanism is more consistent with empirical data that suggests that the level of co-evolution necessary to drive co-speciation of host and pathogen is rarely encountered in nature (Vienne et al., 2013). As models of molecular adaptation and gene tree evolution improve, we may be able to identify phylogenetic congruence at the gene-tree level or signatures of selection that co-vary among hosts and parasites with more confidence. In turn, we might be better able to interpret results in the light of co-diversification vs co-speciation of studies, such as by Tso *et al.* (2018), where a pathogenic strain of *Candidas albicans* evolved into a gut symbiont in mice in just ten weeks. Parallel genomic analyses of the pathogen showed that genes involved with an important virulence factor in *C. albicans*, the hyphal morphogenesis program, had undergone rapid degeneration via both point mutations and deletions.

From candidate genes to whole genome analysis

As outlined above, the co-evolution of host and pathogens can result in distinct and measurable genomic signatures of selection, which can reveal the genetic mechanisms by which hosts and

their pathogens interact. The genomics revolution has spurred the transition from single-candidate gene studies to genome wide analyses of hosts and pathogens. Historically, a number of different candidate host immune genes families have attracted particular attention for studies of host-pathogen interactions, including components of the innate immune system such as toll-like receptors (TLRs) (Tschirren et al., 2013; Zhang et al., 2014; Zhang, Lun & Tsui, 2015; Shan et al., 2018), interferons and antimicrobial peptides (Clark & Wang, 1997; Tennessen, 2005; Franzenburg et al., 2013; Carlin et al., 2018). These and other studies show, both in vertebrates and insects, widespread signatures of positive selection and rapid evolution in genes of the innate immune system (Świderská et al., 2018; Harpur et al., 2019; Adrian et al., 2019). Gene expression studies have also revealed widespread activation of host innate immune genes upon natural or experimental infection with pathogens, such as *Pseudomonas* and *Daphnia* (Kumar et al., 2018). As such, these studies have contributed much to our general understanding of the host's responses to pathogen exposure and common pathways to resistance evolution over time.

The candidate gene family that has attracted the most attention in ecological and evolutionary host-pathogen studies, at least in vertebrates, is the major histocompatibility complex (MHC) (Hughes & Nei, 1988; Bernatchez & Landry, 2003; Meyer-Lucht & Sommer, 2005; Spurgin & Richardson, 2010). MHC genes encode cell-surface molecules that play a central role in pathogen recognition as part of the adaptive immune response. T-cells act to destroy infected cells both directly, as cytotoxic T-cells, and indirectly, as T-helper cells which activate other immune cells, but they can only determine what is self or foreign from peptides presented by MHC molecules. The number of MHC gene copies carried by individuals varies widely between, and even within, species (Kelley, Walter & Trowsdale, 2004; Cheng et al., 2012; Lighten et al., 2014). Additionally, the allelic diversity recorded within and between gene copies makes the MHC genes the most polymorphic loci to date (Reche & Reinherz, 2003; Robinson et al., 2015). This exceptional polymorphism is believed to be primarily maintained by selection from a wide range of pathogens (Prugnolle et al., 2005; Qutob et al., 2011). Overall, it is clear that MHC genes play a pivotal role in the fight against pathogens and numerous studies have established associations between MHC genotypes and infections with a particular pathogen (Kaslow et al., 1996; Meyer-Lucht & Sommer, 2005; Oliver, Telfer & Piertney, 2009; Bolnick & Stutz, 2017).

Although the candidate gene approach has been the standard method for studying immune genes in the context of host-pathogen interactions, the rapidly decreasing costs of high throughput sequencing are making whole-genome approaches much more feasible. Whole-genome population genetic and comparative genomic studies allow unbiased detection of regions of the genome that are evolving non-neutrally across a variety of time scales. When combined with functional annotations or association studies, such comprehensive genome-wide surveys permit incisive tests of the effects of pathogens on host genomes that are simply not possible from candidate gene studies. Furthermore, whole genome studies are not restricted by *a priori* predictions of which genes are important in responding to pathogen challenges. Thus, whole-genome approaches offer the potential to reveal new unbiased insights into the genetic basis of host-pathogen interactions, e.g. (Enard et al., 2016; Shultz & Sackton, 2019). Since multiple genes are most often involved in a host's response to a particular pathogen, whole-genome approaches also have the potential to reveal these understudied polygenic responses (Daub et al., 2013). However, a major caveat associated with the whole-genome approach is that genomic regions of high repeat content, or highly duplicated genes, often do not assemble well or at all, whether considering host or pathogen. Genome assembly problems may be a significant disadvantage for host-pathogen studies given that some key genes which play a role in innate and adaptive immunity are not only highly duplicated but also to some degree physically linked in the genome, such as the beta-defensin and MHC genes (Kaufman et al., 1999; Hellgren & Ekblom, 2010; Balakrishnan et al., 2010). Improved sequence and scaffolding techniques are being developed to remedy problems of assembling such regions (Dilthey et al., 2015) and some may be possible to overcome with long-read sequencing methods. Thus, while the whole-genome approach may become a gold standard for many host-pathogen studies in the future, we currently see a continued need for sequencing methods that target focal genes, in particular in studies of non-model species.

Genomic detection and surveys of pathogens

Ever since the invention of the polymerase chain reaction, molecular approaches have continuously provided sensitive methods for the detection of pathogens, often without prior separation from the host tissues (e.g. malaria pathogens Snounou et al., 1993; Hellgren, Waldenström & Bensch, 2004). High throughput sequencing techniques have now become pivotal for both detection and identification of new pathogens, especially in cases of emerging infectious diseases, and in pathogens with complex life histories and co-infections (Blasco-Costa & Poulin, 2017). Furthermore, unmapped reads in host genome projects are likely a

fruitful source of undiscovered pathogens (Laine et al., 2019). Improved pathogen genomics holds great potential to advance our current understanding of host-pathogen interactions in several ways: From an epidemiological perspective, it allows one to reconstruct the spatial spread of pathogen invasion, illuminates pathogen population dynamics, and enables forecasting of future infection scenarios. Although this has been possible previously by using only a few key genetic markers from samples that spanned decades in time, such as in studies of influenza or rabies virus (Biek et al., 2007; Streicker et al., 2016), whole genome information now allows for high-resolution characterization of outbreaks over shorter timescales (e.g. Ebola (Dudas et al., 2017) and Zika (Faria et al., 2017; Grubaugh et al., 2017)). In addition, open source genomic data-sharing sites and analysis platforms like nextstrain (http://www.nextstrain.org/) are invaluable to explore pathogen time-space variation in real-time. Moreover, genomic analyses of dated pathogen samples have proven successful in inferring directionality of pathogen spread, for example, among wildlife and livestock, thus informing effective control measures (Kamath et al., 2016).

However, many technical challenges still remain for such approaches, especially in situations where pathogens cannot be physically separated from hosts (see Box 1 for an example). For example, pathogen DNA typically makes up only a small fraction of the total extracted DNA from samples of infected hosts, and host samples must therefore be sequenced at >1000X to obtain a 10X coverage of the pathogen (Videvall, 2019). A large number of enrichment protocols for high-throughput sequencing methods have been developed to facilitate the detection and quantification of pathogens. These enrichment protocols are often efficient ways of increasing the ratio of pathogen to host DNA. Before DNA extraction, intracellular pathogens can sometimes be isolated from infected host cells using cell-sorting or laser-capture microscopy techniques (Saliba et al., 2014; Wang et al., 2015), or separated from the host cells by targeting different life stages (e.g. gametes, spores) (Palinauskas et al., 2013). After DNA extraction, selective whole-genome amplification can specifically enrich for pathogen sequences in various ways: (i) by using oligos that are more abundant in the pathogen genome (Melnikov et al., 2011); (ii) by targeting differences in methylation between host and pathogen genomes (Gómez-Díaz et al., 2012); or (iii) by sequence capture enrichment protocols for pathogen DNA (Tagle et al., 1993). When enrichment protocols are not feasible, host and pathogen associated reads can often be separated in silico using reference sequence databases. In such cases, low-coverage detection of genome fragments of pathogens in host genome sequencing reads is a straightforward and fruitful approach (e.g., Laine et al. 2019). Using this approach, putative RNA

viruses of *Drosophila melanogaster* were identified from de novo assembled RNAseq reads (Webster et al., 2015). Dual sequencing analysis of both host and pathogen can be further exploited to characterize the physiological response throughout the course of an infection (Florens et al., 2002; Jean Beltran et al., 2017). However, sequencing coverage and cost are major factors determining feasibility and scope of a study. Enrichment and optimization of protocols carry the caveat that they are study specific and, in many cases, not universally applicable.

Simultaneous genome sequencing of multiple species – metagenomics - can help the field expand beyond the two-organism framework (Westermann, Barquist & Vogel, 2017), as has been most extensively demonstrated in microbiome research in the context of host health (Sekirov et al., 2010; Kamada et al., 2013). It is now clear that the whole microbial community shape host health, but are also in turn selected for and manipulated by hosts (Näpflin & Schmid-Hempel, 2016; Schwarz, Moran & Evans, 2016; Rolhion & Chassaing, 2016; Näpflin & Schmid-Hempel, 2017). In particular, metagenomics is increasingly able to shed light on the function of individual members of the microbiome, for example, by investigating metabolic pathways present in the community (Lee & Hase, 2014). Similarly, sophisticated pathogen-specific screening tools such as sequence chips with known pathogen probes can effectively screen complex ecosystems for pathogens within the community and may identify potential disease reservoirs (Bird & Mazet, 2018). Such approaches are employed by the PREDICT project of USAID which attempts to identify new zoonotic threats in “hot spot” regions in Africa, Asia, and Latin America by sampling wildlife (particularly non-human primates, bats, and rodents) as well as people with close contact with wildlife (http://www.vetmed.ucdavis.edu/ohi/predict/).

Overall, genome-wide techniques and approaches provide us with an unprecedented wealth of information upon which specific hypotheses can be formulated and experimentally tested. A lingering limitation to the impact of such studies is low quality and poor annotation of reference genomes, especially for non-model host species. This challenge considerably slows our rate of discovery because many important parts of the host genome that respond to pathogen infection may remain undiscovered if they do not assemble properly or lack known gene annotation. Furthermore, relevant links to host-pathogen interaction could be missed because the link between genetics and the expressed phenotype is only poorly understood (e.g. the layer of mucus covering the mucosal surface in vertebrates whose composition is relevant for the specific host-pathogen interaction; see Box 2).

Infrastructural challenges of generalizing results across species and systems

The genomic data revolution driven by high-throughput sequencing has created numerous exciting opportunities to study host-pathogen interactions in a multitude of systems in unprecedented detail. This revolution extends to non-model organisms, although inference here may be hampered by a lack of suitable and/or sufficient host or pathogen samples. Moreover, even when such obstacles can be overcome, two major problems of relevance to this review continue to constrain the full impact, reuse, synthesis and reproducibility of host-pathogen studies, particularly for non-model systems: both involve the deposition and analysis of associated sequence data. First, while it is standard practice for sequence data to be deposited in well-curated, detail-rich national databases such as the National Center for Biotechnology Information (NCBI), the European Nucleotide Archive (ENA), or the DNA Data Bank of Japan (DDBJ), the associated biological metadata of these samples are often inadequate or misleading due to various inconsistencies in available sample information that is being deposited. This problem is not trivial to resolve in the context of host-pathogen studies, in part because the complexity of the standardized metadata forms of these platforms for deposition and retrieval of data (Dugan et al., 2014; Chang et al., 2016; Singh et al., 2019). Second, the analysis of genomic data is preceded by a large number of computationally complicated pre-processing steps. The choice of algorithm and parameters in this pre-processing procedure can often have significant impacts on the final results but are generally inadequately documented and communicated. Together, the missing metadata and the lack of transparency regarding computational tools confound or even prevent robust meta-analysis and comparative studies; and without meta-analyses and comparative studies, results from individual studies of various host-pathogen systems cannot be integrated into a larger context.

Improving the availability of metadata and the transparency of computational tools requires the compliance and openness of researchers to collect and analyze their data in a standardized format, with the goal of making the data useable for comparative studies and ultimately to make the data publicly available. Because the interactions between hosts and their pathogens are inherently dynamic across space and time, accurate information on sampling location and timing is essential information to include in metadata. This includes the host source of isolation for pathogens and the infection prevalence of hosts. Importantly, the nomenclature of genetically identical strains or species must be consistent. For example, despite being

genetically identical the haemosporidian lineage "*Haemoproteus* lineage 22" from birds, first described in 2002, has repeatedly been named differently in publications appearing between 2002-2009: "AP21", "COLL2", "SWTH.H2", and "WHA24" (Bensch, Hellgren & Pérez-Tris, 2009). An obvious first step is to improve the design, user-friendliness and programming interface of currently existing popular databases for metadata. An integration of a large amount of data sources has been developed for some systems, such as haemosporidian pathogens in birds (MalAvi database, (Bensch, Hellgren & Pérez-Tris, 2009), or influenza viruses (GISAID database, Yuelong Shu, 2017). Such efforts should ideally be extended to all host-pathogen systems and are being realized more systematically under the umbrella of The Eukaryotic Pathogen Genomics Resource (EuPathDB), a database of pathogen genomic data that currently includes a dozen pathogen groups (Aurrecoechea et al., 2017).

Similar to metadata documentation, detailed documentation of analysis tools and parameter choice is becoming more widely advocated (Nature Editorial, 2017). The use of scripted pipelines and version-controlled analyses has been advocated to address these challenges (e.g. Nunez-Iglesias, 2015). At the most basic level this includes a scripted analysis that does not require manual command input and thus is completely repeatable given the same raw data (and sufficient computational time) (Beaulieu-Jones & Greene, 2016). Hence, depositing analysis scripts in repositories such as Dryad or Github will become an important component of comparative biology in the genomic era in general and in particular of host-pathogens interaction studies. At its best, automated and curated pipelines that allow continuous reanalysis of new and existing data will become an ambitious future goal for comparative studies (e.g. The Lair (pachterlab.github.io/lair/about/)). Although such practices will improve reproducibility of studies and integration of results across studies, such improvement does not necessarily ensure and validate choice of appropriate methods (Lotterhos, Moore & Stapleton, 2018). Nevertheless, extending automated analysis to organisms with more limited genomic resources, which might permit linking of metadata (such as whether a study is experimental or naturally observed) with sequence data across studies would open up exciting frontiers in comparative studies of host-pathogen interactions across different systems and beyond.

Studying host-pathogen interactions in the Anthropocene

The number of pathogen infections are predicted to continue to increase in the near future, as climate change, human population growth and transportation impact the geographic distribution

and contact rate of hosts and pathogens (Altizer, Bartel & Han, 2011; Maganga et al., 2014; Snäll et al., 2015). It has been estimated that wildlife is the source for 72% of emerging infectious diseases in humans (Cleaveland, Laurenson & Taylor, 2001; Jones et al., 2008; Olival et al., 2017) with recent examples including SARS, a virus in bats and small mammals; the avian influenza type H5N1; and Ebola, originally a virus in fruit bats, which recently caused a human catastrophe in western Africa (Dudas et al., 2017). Obviously, such pathogens can have wide-ranging consequences on global societal stability and economy, and can have devastating effects on natural populations (Daszak, Cunningham & Hyatt, 2000; Sachs & Malaney, 2002; Bonds et al., 2009). In this context, rapid DNA sequencing technologies offer great promise for our understanding of host-pathogen dynamics, and hence the ability to predict and control disease epidemics (Wohl, Schaffner & Sabeti, 2016; Takahashi et al., 2018).

Natural systems are increasingly subjected to anthropogenic stressors, including climate change, urban development, overexploitation, pollution, noise, and transport (Gerber et al., 2014). In recognizing that no host-pathogen system exists in isolation, it is essential to understand how such stressors affect the host's fitness, immune system and pathogen susceptibility. For instance, immunotoxic contaminants can have substantial population level effects by contributing to anthropogenic stress and infectious disease outbreaks (Desforges et al., 2016). This is particularly true for marine and terrestrial top-predators, which, due to their life-history and placement at the top of the food chain, accumulate high levels of ecotoxins. Indeed, high tissue concentrations of persistent pollutants in Baltic seals in the 1970-80s were associated with oviduct occlusions and impaired immune system, leading to sterility and repeated infections (Bergman & Olsson, 1986), and recent work suggest that the same may be true for a wide range of European dolphins and killer whales (Jepson et al., 2016). Such increased levels of ecotoxins may explain the increasing prevalence and severity of diseases in marine wildlife. A detailed understanding of the role of these and other stressors in host-pathogen systems will require multispecies and multi-methodological approaches integrating information at all levels of the system, including trophic interactions, resource availability, life-history and population dynamics, as well as gene expression and selection.

Human intervention also has the potential to alter pathogen communities directly, both by eliminating and by introducing pathogens (Daszak, Cunningham & Hyatt, 2000). Pathogens can play crucial roles as ecosystem engineers (Thomas et al., 1999; Wood & Johnson, 2015). Often, we lack the knowledge to accurately predict how the elimination of one pathogen will

545 affect the host population, other pathogens within the same host population, and their effect on
546 the ecological community (Rogalski et al., 2017). For example, the introduction of invasive
547 species often inadvertently results in the introduction of novel pathogens against which native
548 hosts may possess little or no protection (van Riper et al., 1986). Here again, major future
549 challenges include sample availability, ecological monitoring, and the collection and deposition
550 of appropriate metadata. Additionally, cross-disciplinary scientific integration and communication
551 between scientists, managers and decision-makers are crucial in order to advance global
552 health.
553

554 ## Conclusions and prospects

555 Innovations in genomic techniques have the potential to bring a synthesis to the study of host-
556 pathogen interactions across systems and environmental conditions. We highlighted several
557 recent trends in this perspective for genomic studies of host-pathogen systems: (i) evolutionary
558 genomics approaches have allowed the field to move from a candidate gene approach to
559 investigations at the scale of whole genomes; (ii) the use of genomics for the detection and
560 surveillance of host-pathogen systems; (iii) the challenges of the integrating natural history and
561 ecological metadata and genomic data across systems and timescales due to infrastructural
562 challenges of database integration and transparency; and (iv) the impact of anthropogenic
563 stressors on host-pathogen systems that have consequences for global health. Additionally, our
564 survey of the recent literature of ecological genomics of host-pathogen interactions revealed
565 that studies with spatially and ecologically complex settings are rare, as are detailed studies of
566 host genomic responses to pathogens. Any single host-pathogen study is constrained by limited
567 resources or genomic tractability, the geographical and evolutionary time scales involved as well
568 as environmental complexities. Accordingly, transparent and open science will help to achieve a
569 comprehensive understanding of host-pathogen interactions in general. This will contribute to
570 the integration of findings across the different scales (Fig. 1). A large repertoire of comparable
571 and inter-communicative studies will facilitate a more generalizable understanding of the causes
572 and consequences of host-pathogen interactions and a clearer roadmap to combating the
573 continuous threat of pathogens in a changing world.
574

575 # Boxes:

Box 1: Demonstrating pathogen-induced fitness costs in the wild

Determining where an organism lies on the mutualist-pathogen continuum requires an assessment of the fitness costs (i.e. virulence) elicited by a putative pathogen when it has established itself within its host in its natural habitat. In such a scenario, the feasibility of estimating fitness costs strongly depends, on the one hand, on the magnitude of the fitness effect and, on the other hand, the sample size. For example, severe negative fitness effects in birds due to infections by the introduced malaria pathogen *Plasmodium relictum*, have been readily demonstrated in several species of Hawaiian honeycreepers (*Drepanididae*) (van Riper et al., 1986). However, when *P. relictum* infects host species with which it has presumably co-evolved, observed fitness costs are lower (Bensch et al., 2007). Thus, hypothetically, demonstrating a negative fitness effect (i.e., mortality) of 5% year-on-year in natural populations (assuming a pathogen prevalence of 20% and an annual background mortality of 50%) requires a sample size of more than 2000 host individuals and the ability to accurately measure individual survival. The situation becomes even more complex when hosts are repeatedly exposed to the same pathogen and mortality varies across exposures. For example, if mortality is highest upon primary infection, in year two individuals that were unexposed in year one will be at a higher risk of dying than individuals that have been previously exposed. Often only long-term studies, such as that conducted by Asghar et al., (2014) on the effects of *Plasmodium* on lifetime fitness and survival of Great Reed Warblers, provide the sensitivity required to detect fitness costs.

Given that the ecological role of an organism can be dynamic, the fitness consequences for a host of a particular pathogen are strongly dependent on the environmental and genetic context. The most obvious illustration of this is variation in virulence associated with host switches: *Mycoplasma* infection in house finches as compared to other song birds (Ley et al., 2016), Ebola virus in humans as compared to bats (Leroy et al., 2005), or the morbillivirus Phocine Distemper Virus (PDV) in harbor seals as compared to other Arctic pinniped species (Härkönen et al., 2006) are all cases where virulence dramatically increased after switching to a new host. Second, another level of complexity presents itself in the cases of complex pathogen life cycles, where pathogens may require multiple host species for different developmental stages in order to complete their life cycle (Parker et al., 2003; Blasco-Costa &

Poulin, 2017). In cases such as these, it is often difficult to differentiate between pathogen species and different pathogen life stages morphologically. Third, infections by a single pathogen might actually be rare in nature, instead co-infections by multiple pathogen strains or species are likely to be the norm (Petney & Andrews, 1998). In this context, exposure history and the timing of infection might play crucial roles in terms of host fitness and pathogen virulence (Telfer et al., 2010; Ben-Ami, Rigaud & Ebert, 2011). Fourth, pathogen prevalence may vary across space and time and, hence, these patterns need to be taken into consideration in comparisons across scales (Thompson, 2009). This can be on a small scale within a host (e.g. between tissues), or across geographical space (e.g. between populations/species). For example, comparison of host and viral population structure suggests that dispersing male bats spread the rabies virus between genetically isolated female populations (Streicker et al., 2016). Fifth, hosts and their pathogens rarely interact in isolation but rather as part of a larger ecosystem, which might modulate how a pathogen interacts with its host and vice versa (Graham, 2008).

Overall, the availability of large genomic datasets has been pivotal in untangling each of the five levels of complexity. Nevertheless, relying solely on genetic data can be misleading. While new techniques help to identify new pathogens, ecological patterns, and link the genetic structure of host and pathogen populations, the resulting data are ultimately correlational and cannot establish any causal relationships without an experimental approach. For example, sticklebacks (*Gasterosteus aculeatus*) caught in a lake harbored more macroparasites than those from a river (Wegner, Reusch & Kalbe, 2003). With only this observation, one might be tempted to conclude that the sticklebacks from lakes were more susceptible to parasitism than those from rivers. However, subsequent experiments revealed that sticklebacks from lakes were less susceptible to pathogens, but probably experienced higher pathogen exposure (Scharsack & Kalbe, 2014). This illustrates the need for experimental studies to confirm causal relationships implicated by field data. However, experiments are restricted in the complexity they can represent (Plowright et al., 2008). In conclusion, the interpretation of genetic data without a deep understanding of the host-pathogen ecology, and vice versa, can be misleading.

576
577

Box 2: Barriers to infections – an example of difficulties linking genotype and phenotype

Hosts are continuously exposed to potential pathogens, yet the establishment of an infection upon encounter is a relatively rare event. Most pathogenic infections are successfully prevented by "simple" barriers, the host's first lines of defense (McGuckin et al., 2011; Hall, Bento & Ebert, 2017). One of the most underappreciated pre-infection barriers in (non-human) ecology is the continuously secreted layer of mucus covering the mucosal surface in vertebrates (Fig. 3A), and the glycocalyx that covers other epithelial cells and surrounds some single celled organisms (Quintana-Hayashi et al., 2018). As opposed to the skin, which is a dry, acidic, and of much smaller surface, the mucosal surface is orders of magnitude larger and presents a semipermeable, humid environment that many bacteria and pathogens could thrive in. However, the mucosal surfaces are protected by several layers of defense that a pathogen has to circumvent to either gain access to close interactions with host cells, or entry into host cells, or transferring across the host epithelium. The first barrier the pathogens encounter is the continuously secreted mucus layer covering the cells and the epithelial glycocalyx (Quintana-Hayashi et al., 2018), into which a range of antimicrobial molecules are secreted, the bulk of this layer consists of a massive amount of highly diverse glycans (Fig. 3B). Among these highly diverse glycoconjugates, there are those that act as protection against infection by binding and disseminating the pathogen, act as steric hindrance or releasable decoys, but also those that act as receptors for pathogens and confer intimate adherence (Linden et al., 2008; Lindén et al., 2009; Padra et al., 2018).

In fact, across mammalian and teleost species, most known interactions between viral or bacterial pathogens and its host occur via host glycan structures (Aspholm-Hurtig, 2004; Linden et al., 2008; Venkatakrishnan, Packer & Thaysen-Andersen, 2013; Padra et al., 2014; Skoog et al., 2017). Interactions between host glycans and pathogens are thus central for host-pathogen specificity and virulence. As such, one would expect that host glycans and pathogen adhesins are subjected to strong selective pressure (Lindén et al., 2008; Lind n et al., 2010; Vitiazeva et al., 2015; Venkatakrishnan et al., 2017; 2019). While certain individual interactions between host glycans and pathogen adhesins have been dissected in detail (Rydell et al., 2011; Bugaytsova et al., 2017) it remains difficult to actually identify different glycoconjugate compositions and their underlying genetic basis.

While enzymes involved in glycan biosynthesis are easily identified based on sequence identity (curated collection: www.cazy.org; (Lombard et al., 2013) and make up about 5% of the total genome (Rini, Varki & Esko, 2015) the resulting glycan structures are governed by stochastic events, substrate availability and state of differentiation and physiological environment. Thus, with the currently available knowledge it is not feasible to predict glycan repertoire and biosynthetic machinery based solely on genomic and/or transcriptomic sequence data of the host. In addition, we currently lack ability to screen large sample sets for glycan repertoire because mass spectrometric based glycomics discovery is at its best only semi-automatic. Additionally, on the pathogen side, most adhesins of pathogenic organisms have yet to be identified and characterized.

578

Acknowledgment:

We thank Sinan Sharba (University of Gothenburg), for the photographs in figure 3.

Funding Statement:

The themed activities of “Origins of Biodiversity”, of which our host-pathogen workshop was a part, were funded by Chalmers University of Technology and the University of Gothenburg. This work was supported by the Swedish Research Council Formas, the Swedish Research Council, the Alexander von Humboldt Foundation, Kungliga Fysiografiska Sällskapet i Lund, the Carl Tryggers Foundation (postdoctoral fellowship to VAE), and the Swiss National Science Foundation (postdoctoral fellowships nr.168911and 180862 to KN). No additional external funding was received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Adrian J, Bonsignore P, Hammer S, Frickey T, Hauck CR 2019. Adaptation to host-specific bacterial pathogens drives rapid evolution of a human innate immune receptor. *Current Biology* 29:616–630.e5. DOI: 10.1016/j.cub.2019.01.058.

Altizer S, Bartel R, Han BA 2011. Animal migration and infectious disease risk. *Science* 331:296–302. DOI: 10.1126/science.1194694.

Asghar M, Hasselquist D, Hansson B, Zehtindjiev P, Westerdahl H, Bensch S 2015. Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds. *Science* 347:436–438. DOI: 10.1126/science.1261121.

Aspholm-Hurtig M 2004. Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* 305:519–522. DOI: 10.1126/science.1098801.

Aurrecoechea C, Barreto A, Basenko EY, Brestelli J, Brunk BP, Cade S, Crouch K, Doherty R, Falke D, Fischer S, Gajria B, Harb OS, Heiges M, Hertz-Fowler C, Hu S, Iodice J, Kissinger JC, Lawrence C, Li W, Pinney DF, Pulman JA, Roos DS, Shanmugasundram A, Silva-Franco F, Steinbiss S, Stoeckert CJ, Spruill D, Wang H, Warrenfeltz S, Zheng J 2017. EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Research* 45:D581–D591. DOI: 10.1093/nar/gkw1105.

Balakrishnan CN, Ekblom R, Völker M, Westerdahl H, Godinez R, Kotkiewicz H, Burt DW, Graves T, Griffin DK, Warren WC, Edwards SV 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC biology* 8:29. DOI: 10.1186/1741-7007-8-29.

Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, John M, Haas DW, Martinez-Picado J, Dalmau J, López-Galíndez C, Casado C, Rauch A, Günthard HF, Bernasconi E, Vernazza P, Klimkait T, Yerly S, O'Brien SJ, Listgarten J, Pfeifer N, Lippert C, Fusi N, Kutalik Z, Allen TM, Müller V, Harrigan PR, Heckerman D, Telenti A, Fellay J, for the HIV Genome-to-Genome Study and the Swiss HIV Cohort Study 2013. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife* 2:e1001123. DOI: 10.7554/eLife.01123.

Beaulieu-Jones BK, Greene CS 2016. Reproducible computational workflows with continuous analysis. *bioRxiv*:056473. DOI: 10.1101/056473.

Ben-Ami F, Rigaud T, Ebert D 2011. The expression of virulence during double infections by different parasites with conflicting host exploitation and transmission strategies. *Journal of Evolutionary Biology* 24:1307–1316. DOI: 10.1111/j.1420-9101.2011.02264.x.

Benjamini Y, Hochberg Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. DOI: 10.2307/2346101.

Bensch S, Hellgren O, Pérez-Tris J 2009. MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Molecular Ecology Resources* 9:1353–1358. DOI: 10.1111/j.1755-0998.2009.02692.x.

Bensch S, Waldenström J, Jonzén N, Westerdahl H, Hansson B, Sejberg D, Hasselquist D 2007. Temporal dynamics and diversity of avian malaria parasites in a single host species. *The Journal of Animal Ecology* 76:112–122. DOI: 10.1111/j.1365-2656.2006.01176.x.

634 Bergman A, Olsson M 1986. High frequency of skeletal deformities in skulls of the baltic grey
635 seal. *Marine Mammals Committee* 15.
636 Bernatchez L, Landry C 2003. MHC studies in nonmodel vertebrates: what have we learned
637 about natural selection in 15 years? *Journal of Evolutionary Biology* 16:363–377. DOI:
638 10.1046/j.1420-9101.2003.00531.x.
639 Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA 2007. A high-resolution genetic
640 signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of*
641 *the National Academy of Sciences* 104:7993–7998. DOI: 10.1073/pnas.0700741104.
642 Bird BH, Mazet JAK 2018. Detection of emerging zoonotic pathogens: an integrated one health
643 approach. *Annual Review of Animal Biosciences* 6:121–139. DOI: 10.1146/annurev-animal-
644 030117-014628.
645 Biswas S, Akey JM 2006. Genomic insights into positive selection. *Trends in Genetics* 22:437–
646 446.
647 Blasco-Costa I, Poulin R 2017. Parasite life-cycle studies: a plea to resurrect an old
648 parasitological tradition. *Journal of Helminthology* 9:1–10. DOI:
649 10.1017/S0022149X16000924.
650 Bolnick DI, Stutz WE 2017. Frequency dependence limits divergent evolution by favouring rare
651 immigrants over residents. *Nature* 546:285–288. DOI: 10.1038/nature22351.
652 Bonds MH, Keenan DC, Rohani P, Sachs JD 2009. Poverty trap formed by the ecology of
653 infectious diseases. *Proceedings of the Royal Society of London B: Biological Sciences*
654 277:rspb20091778–1192. DOI: 10.1098/rspb.2009.1778.
655 Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB,
656 DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJD, Herring DA, Bauer P, Poinar HN,
657 Krause J 2011. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature*
658 478:506–510. DOI: 10.1038/nature10549.
659 Bourgeois Y, Roulin AC, Müller K, Ebert D 2017. Parasitism drives host genome evolution:
660 Insights from the *Pasteuria ramosa*-Daphnia magna system. *Evolution* 71:1106–1113. DOI:
661 10.1111/evo.13209.
662 Bugaytsova JA, Björnham O, Chernov YA, Gideonsson P, Henriksson S, Mendez M, Sjöström
663 R, Mahdavi J, Shevtsova A, Ilver D, Moonens K, Quintana-Hayashi MP, Moskalenko R,
664 Aisenbrey C, Bylund G, Schmidt A, Åberg A, Brännström K, Königer V, Vikström S,
665 Rakhimova L, Hofer A, Ögren J, Liu H, Goldman MD, Whitmire JM, Ådén J, Younson J,
666 Kelly CG, Gilman RH, Chowdhury A, Mukhopadhyay AK, Nair GB, Papadakos KS,
667 Martinez-Gonzalez B, Sgouras DN, Engstrand L, Unemo M, Danielsson D, Suerbaum S,
668 Oscarson S, Morozova-Roche LA, Olofsson A, Gröbner G, Holgersson J, Esberg A,
669 Strömberg N, Landström M, Eldridge AM, Chromy BA, Hansen LM, Solnick JV, Lindén SK,
670 Haas R, Dubois A, Merrell DS, Schedin S, Remaut H, Arnqvist A, Berg DE, Borén T 2017.
671 *Helicobacter pylori* adapts to chronic infection and gastric disease via pH-responsive BabA-
672 mediated adherence. *Cell Host & Microbe* 21:376–389. DOI: 10.1016/j.chom.2017.02.013.
673 Cairns J, Frickel J, Jalasvuori M, Hiltunen T, Becks L 2017. Genomic evolution of bacterial
674 populations under coselection by antibiotics and phage. *Molecular Ecology* 26:1848–1859.
675 DOI: 10.1111/mec.13950.
676 Carlin AF, Vizcarra EA, Branche E, Viramontes KM, Suarez-Amaran L, Ley K, Heinz S, Benner
677 C, Shresta S, Glass CK 2018. Deconvolution of pro- and antiviral genomic responses in

678 Zika virus-infected and bystander macrophages. *Proceedings of the National Academy of*
679 *Sciences* 115:E9172–E9181. DOI: 10.1073/pnas.1807690115.
680 Chang WE, Peterson MW, Garay CD, Korves T 2016. Pathogen metadata platform: software for
681 accessing and analyzing pathogen strain information. *BMC Bioinformatics* 17:379. DOI:
682 10.1186/s12859-016-1231-2.
683 Cheng Y, Stuart A, Morris K, Taylor R, Siddle H, Deakin J, Jones M, Amemiya CT, Belov K
684 2012. Antigen-presenting genes and genomic copy number variations in the Tasmanian
685 devil MHC. *BMC Genomics* 13:87. DOI: 10.1186/1471-2164-13-87.
686 Chomwong S, Charoensapsri W, Amparyup P, Tassanakajon A 2018. Two host gut-derived
687 lactic acid bacteria activate the proPO system and increase resistance to an AHPND-
688 causing strain of *Vibrio parahaemolyticus* in the shrimp *Litopenaeus vannamei*.
689 *Developmental & Comparative Immunology* 89:54–65. DOI: 10.1016/j.dci.2018.08.002.
690 Clark AG, Wang L 1997. Molecular population genetics of *Drosophila* immune system genes.
691 *Genetics* 147:713–724.
692 Cleaveland S, Laurenson MK, Taylor LH 2001. Diseases of humans and their domestic
693 mammals: pathogen characteristics, host range and the risk of emergence. *Philosophical*
694 *Transactions of the Royal Society B: Biological Sciences* 356:991–999.
695 Daszak P, Cunningham AA, Hyatt AD 2000. Emerging infectious diseases of wildlife – threats to
696 biodiversity and human health. *Science* 287:443–449. DOI: 10.1126/science.287.5452.443.
697 Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L
698 2013. Evidence for polygenic adaptation to pathogens in the human genome. *Molecular*
699 *Biology and Evolution* 30:1544–1558. DOI: 10.1093/molbev/mst080.
700 Desforges J-PW, Sonne C, Levin M, Siebert U, De Guise S, Dietz R 2016. Immunotoxic effects
701 of environmental pollutants in marine mammals. *Environment International* 86:126–139.
702 DOI: 10.1016/j.envint.2015.10.007.
703 Dilthey A, Cox C, Iqbal Z, Nelson MR, McVean G 2015. Improved genome inference in the MHC
704 using a population reference graph. *Nature Genetics* 47:682–688. DOI: 10.1038/ng.3257.
705 Dobson A, Lafferty KD, Kuris AM, Hechinger RF, Jetz W 2008. Colloquium paper: homage to
706 Linnaeus: how many parasites? How many hosts? *PNAS* 105 Suppl 1:11482–11489. DOI:
707 10.1073/pnas.0803232105.
708 Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, Park DJ, Ladner JT, Arias A,
709 Asogun D, Bielejec F, Caddy SL, Cotten M, D'Ambrozio J, Dellicour S, Di Caro A, Diclaro
710 JW, Duraffour S, Elmore MJ, Fakoli LS, Faye O, Gilbert ML, Gevao SM, Gire S, Gladden-
711 Young A, Gnirke A, Goba A, Grant DS, Haagmans BL, Hiscox JA, Jah U, Kugelman JR, Di
712 Liu, Lu J, Malboeuf CM, Mate S, Matthews DA, Matranga CB, Meredith LW, Qu J, Quick J,
713 Pas SD, Phan MVT, Pollakis G, Reusken CB, Sanchez-Lockhart M, Schaffner SF,
714 Schieffelin JS, Sealfon RS, Simon-Loriere E, Smits SL, Stoecker K, Thorne L, Tobin EA,
715 Vandi MA, Watson SJ, West K, Whitmer S, Wiley MR, Winnicki SM, Wohl S, Wölfel R,
716 Yozwiak NL, Andersen KG, Blyden SO, Bolay F, Carroll MW, Dahn B, Diallo B, Formenty P,
717 Fraser C, Gao GF, Garry RF, Goodfellow I, Günther S, Happi CT, Holmes EC, Kargbo B,
718 Keïta S, Kellam P, Koopmans MPG, Kuhn JH, Loman NJ, Magassouba N, Naidoo D, Nichol
719 ST, Nyenswah T, Palacios G, Pybus OG, Sabeti PC, Sall A, Ströher U, Wurie I, Suchard
720 MA, Lemey P, Rambaut A 2017. Virus genomes reveal factors that spread and sustained
721 the Ebola epidemic. *Nature* 544:309–315. DOI: 10.1038/nature22040.

722 Dugan VG, Emrich SJ, Giraldo-Calderón GI, Harb OS, Newman RM, Pickett BE, Schriml LM,
723 Stockwell TB, Stoeckert CJ, Sullivan DE, Singh I, Ward DV, Yao A, Zheng J, Barrett T,
724 Birren B, Brinkac L, Bruno VM, Caler E, Chapman S, Collins FH, Cuomo CA, Di Francesco
725 V, Durkin S, Eppinger M, Feldgarden M, Fraser C, Fricke WF, Giovanni M, Henn MR, Hine
726 E, Hotopp JD, Karsch-Mizrachi I, Kissinger JC, Lee EM, Mathur P, Mongodin EF, Murphy
727 CI, Myers G, Neafsey DE, Nelson KE, Nierman WC, Puzak J, Rasko D, Roos DS,
728 Sadzewicz L, Silva JC, Sobral B, Squires RB, Stevens RL, Tallon L, Tettelin H, Wentworth
729 D, White O, Will R, Wortman J, Zhang Y, Scheuermann RH 2014. Standardized metadata
730 for human pathogen/vector genomic sequences. *PLoS ONE* 9:e99979. DOI:
731 10.1371/journal.pone.0099979.
732 Enard D, Petrov DA 2018. Evidence that RNA viruses drove adaptive introgression between
733 Neanderthals and modern humans. *Cell* 175:360–371.e13.
734 Enard D, Le Cai, Gwennap C, Petrov DA 2016. Viruses are a dominant driver of protein
735 adaptation in mammals. *eLife* 5:e12469. DOI: 10.7554/eLife.12469.
736 Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black
737 A, da Costa AC, Franco LC, Silva SP, Wu CH, Raghwani J, Cauchemez S, Plessis du L,
738 Verotti MP, de Oliveira WK, Carmo EH, Coelho GE, Santelli ACFS, Vinhal LC, Henriques
739 CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu CY, Muñoz-
740 Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA, Chieppe AO,
741 Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO, Siqueira IC, de
742 Queiroz MG, de Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D, Moura LC,
743 Dhalia R, França RF, Magalhães T, Marques ET, Jaenisch T, Wallau GL, de Lima MC,
744 Nascimento V, de Cerqueira EM, de Lima MM, Mascarenhas DL, Neto JPM, Levin AS,
745 Tozetto-Mendoza TR, Fonseca SN, Mendes-Correa MC, Milagres FP, Segurado A, Holmes
746 EC, Rambaut A, Bedford T, Nunes MRT, Sabino EC, Alcantara LCJ, Loman NJ, Pybus OG
747 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas.
748 *Nature* 546:406–410. DOI: 10.1038/nature22401.
749 Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N,
750 Sacci JB, Tabb DL, Witney AA, Wolters D, Wu Y, Gardner MJ, Holder AA, Sinden RE,
751 Yates JR, Carucci DJ 2002. A proteomic view of the *Plasmodium falciparum* life cycle.
752 *Nature* 419:520–526. DOI: 10.1038/nature01107.
753 Franzenburg S, Walter J, Kuenzel S, Wang J, Baines JF, Bosch TCG, Fraune S 2013. Distinct
754 antimicrobial peptide expression determines host species-specific bacterial associations.
755 *Proceedings of the National Academy of Sciences of the United States of America*
756 110:E3730–E3738. DOI: 10.1073/pnas.1304960110.
757 Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R 2011. Signatures of
758 environmental genetic adaptation pinpoint pathogens as the main selective pressure
759 through human evolution. *PLoS genetics* 7:e1002355. DOI: 10.1371/journal.pgen.1002355.
760 Gerber LR, Del Mar Mancha-Cisneros M, O'Connor MI, Selig ER 2014. Climate change impacts
761 on connectivity in the ocean: Implications for conservation. *Ecosphere* 5:1–18.
762 Gómez-Díaz E, Jordà M, Peinado MA, Rivero A 2012. Epigenetics of Host–Pathogen
763 Interactions: The Road Ahead and the Road Behind. *PLoS Pathogens* 8:e1003007. DOI:
764 10.1371/journal.ppat.1003007.

765 Graham AL 2008. Ecological rules governing helminth-microparasite coinfection. *PNAS*
766 105:566–570. DOI: 10.1073/pnas.0707221105.
767 Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M,
768 Arkhipova K, Carmichael M, Cruaud C, Dimier C, Domínguez-Huerta G, Ferland J, Kandels
769 S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay J-É, Vik D, Babin
770 M, Bowler C, Culley AI, de Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Wincker
771 P, Sullivan MB, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G, de Vargas C,
772 Follows M, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D, Jaillon O, Kandels-
773 Lewis S, Karp-Boss L, Karsenti E, Not F, Ogata H, Pesant S, Poulton N, Raes J, Sardet C,
774 Speich S, Stemmann L, Sunagawa S 2019. Marine DNA viral macro- and microdiversity
775 from pole to pole. *Cell*. DOI: 10.1016/j.cell.2019.03.040.
776 Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, Wiley MR, White
777 S, Thézé J, Magnani DM, Prieto K, Reyes D, Bingham AM, Paul LM, Robles-Sikisaka R,
778 Oliveira G, Pronty D, Barcellona CM, Metsky HC, Baniecki ML, Barnes KG, Chak B, Freije
779 CA, Gladden-Young A, Gnirke A, Luo C, MacInnis B, Matranga CB, Park DJ, Qu J,
780 Schaffner SF, Tomkins-Tinch C, West KL, Winnicki SM, Wohl S, Yozwiak NL, Quick J,
781 Fauver JR, Khan K, Brent SE, Reiner RC, Lichtenberger PN, Ricciardi MJ, Bailey VK,
782 Watkins DI, Cone MR, Kopp EW, Hogan KN, Cannons AC, Jean R, Monaghan AJ, Garry
783 RF, Loman NJ, Faria NR, Porcelli MC, Vasquez C, Nagle ER, Cummings DAT, Stanek D,
784 Rambaut A, Sanchez-Lockhart M, Sabeti PC, Gillis LD, Michael SF, Bedford T, Pybus OG,
785 Isern S, Palacios G, Andersen KG 2017. Genomic epidemiology reveals multiple
786 introductions of Zika virus into the United States. *Nature* 546:401–405. DOI:
787 10.1038/nature22400.
788 Haasl RJ, Payseur BA 2016. Fifteen years of genomewide scans for selection: trends, lessons
789 and unaddressed genetic sources of complication. *Molecular Ecology* 25:5–23. DOI:
790 10.1111/mec.13339.
791 Hall MD, Bento G, Ebert D 2017. The evolutionary consequences of stepwise infection
792 processes. *Trends in Ecology & Evolution* 0. DOI: 10.1016/j.tree.2017.05.009.
793 Harpur BA, Guarna MM, Huxter E, Higo H, Moon K-M, Hoover SE, Ibrahim A, Melathopoulos
794 AP, Desai S, Currie RW, Pernal SF, Foster LJ, Zayed A 2019. Integrative genomics reveals
795 the genetics and evolution of the honey bee's social immune system. *Genome Biology and*
796 *Evolution*. DOI: 10.1093/gbe/evz018.
797 Härkönen T, Dietz R, Reijnders P, Teilmann J, Harding K, Hall A, Brasseur S, Siebert U,
798 Goodman SJ, Jepson PD, Rasmussen TD, Thompson P 2006. The 1988 and 2002 phocine
799 distemper virus epidemics in European harbour seals. *Diseases of Aquatic Organisms*
800 68:115–130. DOI: 10.3354/dao068115.
801 Hellgren O, Ekblom R 2010. Evolution of a cluster of innate immune genes (beta-defensins)
802 along the ancestral lines of chicken and zebra finch. *Immunome Research* 6:3. DOI:
803 10.1186/1745-7580-6-3.
804 Hellgren O, Waldenström J, Bensch S 2004. A new PCR assay for simultaneous studies of
805 *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. *Journal of Parasitology*
806 90:797–802. DOI: 10.1645/GE-184R1.
807 Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster
808 D, McMichael AJ, Greenwood BM 1991. Common West African HLA antigens are

associated with protection from severe malaria. *Nature* 352:595–600. DOI: 10.1038/352595a0.

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K 2002. A comprehensive review of genetic association studies. *Genetics in Medicine* 4:45–61. DOI: 10.1097/00125817-200203000-00002.

Hughes AL, Nei M 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167–170. DOI: 10.1038/335167a0.

Jean Beltran PM, Federspiel JD, Sheng X, Cristea IM 2017. Proteomics and integrative omic approaches for understanding host–pathogen interactions and infectious diseases. *Molecular Systems Biology* 13:922. DOI: 10.15252/msb.20167062.

Jepson PD, Deaville R, Barber JL, Aguilar À, Borrell A, Murphy S, Barry J, Brownlow A, Barnett J, Berrow S, Cunningham AA, Davison NJ, Doeschate ten M, Esteban R, Ferreira M, Foote AD, Genov T, Giménez J, Loveridge J, Llavona Á, Martin V, Maxwell DL, Papachlimitzou A, Penrose R, Perkins MW, Smith B, de Stephanis R, Tregenza N, Verborgh P, Fernandez A, Law RJ 2016. PCB pollution continues to impact populations of orcas and other dolphins in European waters. *Scientific Reports* 6:18573. DOI: 10.1038/srep18573.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P 2008. Global trends in emerging infectious diseases. *Nature* 451:990–993. DOI: 10.1038/nature06536.

Kamada N, Chen GY, Inohara N, Nunez G 2013. Control of pathogens and pathobionts by the gut microbiota. *Nature Immunology* 14:685–690. DOI: 10.1038/ni.2608.

Kamath PL, Foster JT, Drees KP, Luikart G, Quance C, Anderson NJ, Clarke PR, Cole EK, Drew ML, Edwards WH, Rhyan JC, Treanor JJ, Wallen RL, White PJ, Robbe-Austerman S, Cross PC 2016. Genomics reveals historic and contemporary transmission dynamics of a bacterial disease among wildlife and livestock. *Nature Communications* 7:11448. DOI: 10.1038/ncomms11448.

Karlsson EK, Kwiatkowski DP, Sabeti PC 2014. Natural selection and infectious disease in human populations. *Nature Reviews Genetics* 15:379–393. DOI: 10.1038/nrg3734.

Kaslow RA, Carrington M, Apple R, Park L, Munoz A 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV–1 infection. *Nature medicine* 2:405–411. DOI: 10.1038/nm0496-405.

Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401:923–925. DOI: 10.1038/44856.

Kelley J, Walter L, Trowsdale J 2004. Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56:683–695. DOI: 10.1007/s00251-004-0717-7.

Koonin EV, Makarova KS, Wolf YI 2017. Evolutionary genomics of defense systems in archaea and bacteria. *Annual Review of Microbiology* 71:233–261. DOI: 10.1146/annurev-micro-090816-093830.

Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genetics* 4:e1000144. DOI: 10.1371/journal.pgen.1000144.

Kratochwil CF, Meyer A 2015. Closing the genotype–phenotype gap: Emerging technologies for evolutionary genetics in ecological model vertebrate systems. *Bioessays* 37:213–226. DOI: 10.1002/bies.201400142.

853 Kumar SS, Tandberg JI, Penesyan A, Elbourne LDH, Suarez-Bosche N, Don E, Skadberg E,
854 Fenaroli F, Cole N, Winther-Larsen HC, Paulsen IT 2018. Dual transcriptomics of host-
855 pathogen interaction of cystic fibrosis isolate *Pseudomonas aeruginosa* PASS1 With
856 Zebrafish. *Frontiers in Cellular and Infection Microbiology* 8:1922. DOI:
857 10.3389/fcimb.2018.00406.
858 Kuris AM, Hechinger RF, Shaw JC, Whitney KL, Aguirre-Macedo L, Boch CA, Dobson AP,
859 Dunham EJ, Fredensborg BL, Huspeni TC, Lorda J, Mababa L, Mancini FT, Mora AB,
860 Pickering M, Talhouk NL, Torchin ME, Lafferty KD 2008. Ecosystem energetic implications
861 of parasite and free-living biomass in three estuaries. *Nature* 454:515–518. DOI:
862 10.1038/nature06970.
863 Kwiatkowski DP 2005. How malaria has affected the human genome and what human genetics
864 can teach us about malaria. *The American Journal of Human Genetics* 77:171–192.
865 Laine VN, Gossmann TI, van Oers K, Visser ME, Groenen MAM 2019. Exploring the unmapped
866 DNA and RNA reads in a songbird genome. *BMC Genomics* 20:19. DOI: 10.1186/s12864-
867 018-5378-2.
868 Lee W-J, Hase K 2014. Gut microbiota–generated metabolites in animal health and disease.
869 *Nature Chemical Biology* 10:416–424. DOI: 10.1038/nchembio.1535.
870 Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Délicat A, Paweska JT,
871 Gonzalez J-P, Swanepoel R 2005. Fruit bats as reservoirs of Ebola virus. *Nature* 438:575–
872 576. DOI: 10.1038/438575a.
873 Ley DH, Hawley DM, Geary SJ, Dhondt AA 2016. House Finch (Haemorhous mexicanus)
874 Conjunctivitis, and Mycoplasmaspp. Isolated from North American Wild Birds, 1994–2015.
875 *Journal of wildlife diseases* 52:669–673. DOI: 10.7589/2015-09-244.
876 Libertucci J, Young VB 2019. The role of the microbiota in infectious diseases. *Nature*
877 *Microbiology* 4:35–45. DOI: 10.1038/s41564-018-0278-4.
878 Lighten J, van Oosterhout C, Paterson IG, McMullan M, Bentzen P 2014. Ultra-deep Illumina
879 sequencing accurately identifies MHC class IIb alleles and provides evidence for copy
880 number variation in the guppy (*Poecilia reticulata*). *Molecular Ecology Resources* 14:753–
881 767. DOI: 10.1111/1755-0998.12225.
882 Lind n S, Semino-Mora C, Liu H, Rick J, Dubois A 2010. Role of mucin Lewis status in
883 resistance to *Helicobacter pylori* infection in pediatric patients. *Helicobacter* 15:251–258.
884 DOI: 10.1111/j.1523-5378.2010.00765.x.
885 Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA 2008. Mucins in the mucosal
886 barrier to infection. *Mucosal Immunology* 1:183–197. DOI: 10.1038/mi.2008.5.
887 Lindén SK, Sheng YH, Every AL, Miles KM, Skoog EC, Florin THJ, Sutton P, McGuckin MA
888 2009. MUC1 limits *Helicobacter pylori* Infection both by steric hindrance and by acting as a
889 releasable decoy. *PLoS Pathogens* 5:e1000617. DOI: 10.1371/journal.ppat.1000617.
890 Lindén S, Mahdavi J, Semino-Mora C, Olsen C, Carlstedt I, Borén T, Dubois A 2008. Role of
891 ABO secretor status in mucosal innate immunity and *H. pylori* infection. *PLoS Pathogens*
892 4:e2. DOI: 10.1371/journal.ppat.0040002.
893 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B 2013. The carbohydrate-
894 active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42:D490–D495. DOI:
895 10.1093/nar/gkt1178.

Lotterhos KE, Moore JH, Stapleton AE 2018. Analysis validation has been neglected in the Age of Reproducibility. *PLoS Biology* 16:e3000070.

Padra M, Adamczyk B, Benktander J, Flahou B, Skoog EC, Padra JT, Smet A, Jin C, Ducatelle R, Samuelsson T, Haesebrouck F, Karlsson NG, Teneberg S, Lindén SK 2018. *Helicobacter suis* binding to carbohydrates on human and porcine gastric mucins and glycolipids occurs via two modes. *Virulence* 9:898–918. DOI: 10.1080/21505594.2018.1460979.

Maganga GD, Bourgarel M, Vallo P, Dallo TD, Ngoagouni C, Drexler JF, Drosten C, Nakouné ER, Leroy EM, Morand S 2014. Bat distribution size or shape as determinant of viral richness in African bats. *PLoS ONE* 9:e100172. DOI: 10.1371/journal.pone.0100172.

Malaria Genomic Epidemiology Network, Band G, Rockett KA, Spencer CCA, Kwiatkowski DP 2015. A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* 526:253–257. DOI: 10.1038/nature15390.

McGuckin MA, Lindén SK, Sutton P, Florin TH 2011. Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology* 9:265–278. DOI: 10.1038/nrmicro2538.

Melnikov A, Galinsky K, Rogov P, Fennell T, Tyne D, Russ C, Daniels R, Barnes KG, Bochicchio J, Ndiaye D, Sene PD, Wirth DF, Nusbaum C, Volkman SK, Birren BW, Gnirke A, Neafsey DE 2011. Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biology* 12:R73. DOI: 10.1186/gb-2011-12-8-r73.

Meyer-Lucht Y, Sommer S 2005. MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Molecular Ecology* 14:2233–2243. DOI: 10.1111/j.1365-294X.2005.02557.x.

Näpflin K, Schmid-Hempel P 2016. Immune response and gut microbial community structure in bumblebees after microbiota transplants. *Proceedings of the Royal Society B: Biological Sciences* 283:–. DOI: 10.1098/rspb.2016.0312.

Näpflin K, Schmid-Hempel P 2017. Host effects on microbiota community assembly. *The Journal of Animal Ecology* 473:174–340. DOI: 10.1111/1365-2656.12768.

Nature Editorial 2017. Announcement: Towards greater reproducibility for life-sciences research. *Nature* 546:8–8. DOI: 10.1038/546008a.

Nielsen R 2005. Molecular signatures of natural selection. *Annual review of genetics* 39:197–218. DOI: 10.1146/annurev.genet.39.073003.112420.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky JJ, Adams MD, Cargill M 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* 3:e170. DOI: 10.1371/journal.pbio.0030170.

Nunez-Iglesias, J., 2015 Why scientists should code in the open. *Available at https://ilovesymposia.com/2015/12/26/why-scientists-should-code-in-the-open/* (accessed July 7, 2017).

Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646–650. DOI: 10.1038/nature22975.

Oliver MK, Telfer S, Piertney SB 2009. Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*).

Proceedings of the Royal Society of London B: Biological Sciences 276:1119–1128. DOI: 10.1098/rspb.2008.1525.

Padra JT, Sundh H, Jin C, Karlsson NG, Sundell K, Linden SK 2014. *Aeromonas salmonicida* binds differentially to mucins isolated from skin and intestinal regions of Atlantic salmon in an N-acetylneuraminic acid-dependent manner. *Infection and Immunity* 82:5235–5245. DOI: 10.1128/IAI.01931-14.

Pal C, Maciá MD, Oliver A, Schachar I, Buckling A 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450:1079–1081. DOI: 10.1038/nature06350.

Palinauskas V, Križanauskienė A, Iezhova TA, Bolshakov CV, Jönsson J, Bensch S, Valkiūnas G 2013. A new method for isolation of purified genomic DNA from haemosporidian parasites inhabiting nucleated red blood cells. *Experimental Parasitology* 133:275–280. DOI: 10.1016/j.exppara.2012.12.003.

Parker GA, Chubb JC, Ball MA, Roberts GN 2003. Evolution of complex life cycles in helminth parasites. *Nature* 425:480–484. DOI: 10.1038/nature02012.

Penczykowski RM, Laine A-L, Koskella B 2015. Understanding the ecology and evolution of host-parasite interactions across scales. *Evolutionary Applications* 9:37–52. DOI: 10.1111/eva.12294.

Petney TN, Andrews RH 1998. Multiparasite communities in animals and humans: frequency, structure and pathogenic significance. *International journal for parasitology* 28:377–393. DOI: 10.1016/S0020-7519(97)00189-6.

Plowright RK, Sokolow SH, Gorman ME, Daszak P, Foley JE 2008. Causal inference in disease ecology: investigating ecological drivers of disease emergence. *Frontiers in Ecology and the Environment* 6:420–429. DOI: 10.1890/070086.

Poulin R 2014. Parasite biodiversity revisited: frontiers and constraints. *International journal for parasitology* 44:581–589. DOI: 10.1016/j.ijpara.2014.02.003.

Prugnolle F, Manica A, Charpentier M, Guégan JF, Guernier V, Balloux F 2005. Pathogen-driven selection and worldwide HLA Class I diversity. *Current Biology* 15:1022–1027. DOI: 10.1016/j.cub.2005.04.050.

Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M, Deschamps M, Naffakh N, Duffy D, Coen A, Leroux-Roels G, Clément F, Boland A, Deleuze J-F, Kelso J, Albert ML, Quintana-Murci L 2016. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell* 167:643–656.e17. DOI: 10.1016/j.cell.2016.09.024.

Quintana-Hayashi MP, Padra M, Padra JT, Benktander J, Lindén SK 2018. Mucus-pathogen interactions in the gastrointestinal tract of farmed animals. *Microorganisms* 6:55. DOI: 10.3390/microorganisms6020055.

Qutob N, Balloux F, Raj T, Liu H, Marion de Procé S, Trowsdale J, Manica A 2011. Signatures of historical demography and pathogen richness on MHC class I genes. *Immunogenetics* 64:165–175. DOI: 10.1007/s00251-011-0576-y.

Reche PA, Reinherz EL 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *Journal of Molecular Biology* 331:623–641. DOI: 10.1016/s0022-2836(03)00750-2.

Rini JM, Varki A, Esko JD 2015. Glycosyltransferases and glycan-processing enzymes. In: *Essentials of Glycobiology*. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. DOI: 10.1101/glycobiology.3e.006.

Robinson CJ, Bohannan BJM, Young VB 2010. From structure to function: the ecology of host-associated microbial communities. *Microbiology and molecular biology reviews: MMBR* 74:453–476. DOI: 10.1128/MMBR.00014-10.

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE 2015. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research* 43:D423–D431. DOI: 10.1093/nar/gku1161.

Rogalski MA, Gowler CD, Shaw CL, Hufbauer RA, Duffy MA 2017. Human drivers of ecological and evolutionary dynamics in emerging and disappearing infectious disease systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372:20160043. DOI: 10.1098/rstb.2016.0043.

Rolhion N, Chassaing B 2016. When pathogenic bacteria meet the intestinal microbiota. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:20150504. DOI: 10.1098/rstb.2015.0504.

Rydell GE, Kindberg E, Larson G, Svensson L 2011. Susceptibility to winter vomiting disease: a sweet matter. *Reviews in Medical Virology* 21:370–382. DOI: 10.1002/rmv.704.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES 2006. Positive natural selection in the human lineage. *Science* 312:1614–1620. DOI: 10.1126/science.1124309.

Sachs J, Malaney P 2002. The economic and social burden of malaria. *Nature* 415:680–685. DOI: 10.1038/415680a.

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG 2007. Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics* 39:1461–1468. DOI: 10.1038/ng.2007.60.

Saliba A-E, Westermann AJ, Gorski SA, Vogel J 2014. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Research* 42:8845–8860. DOI: 10.1093/nar/gku555.

Savage AE, Zamudio KR 2011. MHC genotypes associate with resistance to a frog-killing fungus. *PNAS* 108:16705–16710. DOI: 10.1073/pnas.1106893108.

Scanlan PD, Hall AR, Blackshields G, Friman V-P, Davis MR, Goldberg JB, Buckling A 2015. Coevolution with bacteriophages drives genome-wide host evolution and constrains the acquisition of abiotic-beneficial mutations. *Molecular Biology and Evolution* 32:1425–1435. DOI: 10.1093/molbev/msv032.

Scharsack JR, Kalbe M 2014. Differences in susceptibility and immune responses of three-spined sticklebacks (*Gasterosteus aculeatus*) from lake and river ecotypes to sequential infections with the eye fluke *Diplostomum pseudospathaceum*. *Parasites & Vectors* 7:109. DOI: 10.1186/1756-3305-7-109.

Schmid-Hempel P 2011. *Evolutionary parasitology: The integrated study of infections, immunology*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199229482.001.0001.

Schwarz RS, Moran NA, Evans JD 2016. Early gut colonizers shape parasite susceptibility and microbiota composition in honey bee workers. *Proceedings of the National Academy of Sciences* 113:9345–9350. DOI: 10.1073/pnas.1606631113.

1024 Sekirov I, Russell SL, Antunes LCM, Finlay BB 2010. Gut microbiota in health and disease.
1025 *Physiological Reviews* 90:859–904. DOI: 10.1152/physrev.00045.2009.
1026 Shan S, Liu D, Liu R, Zhu Y, Li T, Zhang F, An L, Yang G, Li H 2018. Non-mammalian Toll-like
1027 receptor 18 (Tlr18) recognizes bacterial pathogens in common carp (*Cyprinus carpio L.):*
1028 Indications for a role of participation in the NF-κB signaling pathway. *Fish & Shellfish*
1029 *Immunology* 72:187–198. DOI: 10.1016/j.fsi.2017.09.081.
1030 Shultz AJ, Sackton TB 2019. Immune genes are hotspots of shared positive selection across
1031 birds and mammals. *eLife* 8:1703. DOI: 10.7554/eLife.41815.
1032 Singh I, Kuscuoglu M, Harkins DM, Sutton G, Fouts DE, Nelson KE 2019. OMeta: an ontology-
1033 based, data-driven metadata tracking system. *BMC Bioinformatics* 20:8. DOI:
1034 10.1186/s12859-018-2580-9.
1035 Skoog EC, Padra M, Åberg A, Gideonsson P, Obi I, Quintana-Hayashi MP, Arnqvist A, Lindén
1036 SK 2017. BabA dependent binding of Helicobacter pylori to human gastric mucins cause
1037 aggregation that inhibits proliferation and is regulated via ArsS. *Scientific Reports* 7:40656.
1038 DOI: 10.1038/srep40656.
1039 Snäll T, OHara RB, Ray C, Collinge SK 2015. Climate-driven spatial dynamics of plague among
1040 prairie dog colonies. *The American Naturalist* 171:238–248. DOI: 10.1086/525051.
1041 Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, Thaithong S, Brown KN
1042 1993. High sensitivity of detection of human malaria parasites by the use of nested
1043 polymerase chain reaction. *Molecular and Biochemical Parasitology* 61:315–320. DOI:
1044 10.1016/0166-6851(93)90077-b.
1045 Spurgin LG, Richardson DS 2010. How pathogens drive genetic diversity: MHC, mechanisms
1046 and misunderstandings. *Proceedings of the Royal Society of London B: Biological Sciences*
1047 277:979–988. DOI: 10.1098/rspb.2009.2084.
1048 Spyrou MA, Bos KI, Herbig A, Krause J 2019. Ancient pathogen genomics as an emerging tool
1049 for infectious disease research. *Nature Reviews Genetics* 13:1. DOI: 10.1038/s41576-019-
1050 0119-1.
1051 Streicker DG, Winternitz JC, Satterfield DA, Condori-Condori RE, Broos A, Tello C, Recuenco
1052 S, Velasco-Villa A, Altizer S, Valderrama W 2016. Host-pathogen evolutionary signatures
1053 reveal dynamics and future invasions of vampire bat rabies. *PNAS* 113:10926–10931. DOI:
1054 10.1073/pnas.1606587113.
1055 Świderská Z, Šmídová A, Buchtová L, Bryjová A, Fabiánová A, Munclinger P, Vinkler M 2018.
1056 Avian Toll-like receptor allelic diversity far exceeds human polymorphism: an insight from
1057 domestic chicken breeds. *Scientific Reports* 8:343. DOI: 10.1038/s41598-018-36226-1.
1058 Tagle DA, Swaroop M, Lovett M, Collins FS 1993. Magnetic bead capture of expressed
1059 sequences encoded within large genomic segments. *Nature* 361:751–753. DOI:
1060 10.1038/361751a0.
1061 Takahashi MK, Tan X, Dy AJ, Braff D, Akana RT, Furuta Y, Donghia N, Ananthakrishnan A,
1062 Collins JJ 2018. A low-cost paper-based synthetic biology platform for analyzing gut
1063 microbiota and host biomarkers. *Nature Communications* 9:3347. DOI: 10.1038/s41467-
1064 018-05864-4.
1065 Telfer S, Lambin X, Birtles R, Beldomenico P, Burthe S, Paterson S, Begon M 2010. Species
1066 interactions in a parasite community drive infection risk in a wildlife population. *Science*
1067 330:243–246. DOI: 10.1126/science.1190333.

1068 Tennessen JA 2005. Molecular evolution of animal antimicrobial peptides: widespread moderate
1069 positive selection. *Journal of Evolutionary Biology* 18:1387–1394. DOI: 10.1111/j.1420-
1070 9101.2005.00925.x.
1071 Thomas F, Poulin R, de Meeüs T, Guégan JF, Renaud F 1999. Parasites and ecosystem
1072 engineering: what roles could they play? *Oikos* 84:167. DOI: 10.2307/3546879.
1073 Thompson JN 2009. *The Coevolutionary Process*. University of Chicago Press.
1074 Tschirren B, Andersson M, Scherman K, Westerdahl H, Mittl PRE, Raberg L 2013.
1075 Polymorphisms at the innate immune receptor TLR2 are associated with *Borrelia* infection
1076 in a wild rodent population. *Proceedings of the Royal Society of London B: Biological*
1077 *Sciences* 280:20130364–20130364. DOI: 10.1098/rspb.2013.0364.
1078 Tso GHW, Reales-Calderon JA, Tan ASM, Sem X, Le GTT, Tan TG, Lai GC, Srinivasan KG,
1079 Yurieva M, Liao W, Poidinger M, Zolezzi F, Rancati G, Pavelka N 2018. Experimental
1080 evolution of a fungal pathogen into a gut symbiont. *Science* 362:589–595. DOI:
1081 10.1126/science.aat0537.
1082 van Riper C, van Riper SG, Goff ML, Laird M 1986. The epizootiology and ecological
1083 significance of *Malaria* in Hawaiian land birds. *Ecological Monographs* 56:327–344. DOI:
1084 10.2307/1942550.
1085 Venkatakrishnan V, Packer NH, Thaysen-Andersen M 2013. Host mucin glycosylation plays a
1086 role in bacterial adhesion in lungs of individuals with cystic fibrosis. *Expert review of*
1087 *respiratory medicine* 7:553–576. DOI: 10.1586/17476348.2013.837752.
1088 Venkatakrishnan V, Padra JT, Sundh H, Sundell K, Jin C, Langeland M, Carlberg H, Vidakovic
1089 A, Lundh T, Karlsson NG, Lindén SK 2019. Exploring the Arctic Charr Intestinal Glycome:
1090 Evidence of Increased N-Glycolylneuraminic Acid Levels and Changed Host-Pathogen
1091 Interactions in Response to Inflammation. *Journal of Proteome Research*. DOI:
1092 10.1021/acs.jproteome.8b00973.
1093 Venkatakrishnan V, Quintana-Hayashi MP, Mahu M, Haesebrouck F, Pasmans F, Lindén SK
1094 2017. Brachyspira hyodysenteriaeInfection Regulates Mucin Glycosylation Synthesis
1095 Inducing an Increased Expression of Core-2 O-Glycans in Porcine Colon. *Journal of*
1096 *Proteome Research* 16:1728–1742. DOI: 10.1021/acs.jproteome.7b00002.
1097 Videvall E 2019. Genomic Advances in Avian Malaria Research. *Trends in Parasitology* 35:254–
1098 266. DOI: 10.1016/j.pt.2018.12.005.
1099 de Vienne DM, Giraud T, Shykoff JA 2007. When can host shifts produce congruent host and
1100 parasite phylogenies? A simulation approach. *Journal of Evolutionary Biology* 20:1428–
1101 1438. DOI: 10.1111/j.1420-9101.2007.01340.x.
1102 de Vienne DM, Refrégier G, López Villavicencio M, Tellier A, Hood ME, Giraud T 2013.
1103 Cospeciation vs host-shift speciation: methods for testing, evidence from natural
1104 associations and relation to coevolution. *New Phytologist* 198:347–385. DOI:
1105 10.1111/nph.12150.
1106 Villarino NF, LeCleir GR, Denny JE, Dearth SP, Harding CL, Sloan SS, Gribble JL, Campagna
1107 SR, Wilhelm SW, Schmidt NW 2016. Composition of the gut microbiota modulates the
1108 severity of malaria. *Proceedings of the National Academy of Sciences of the United States*
1109 *of America*:201504887. DOI: 10.1073/pnas.1504887113.
1110 Vitiazeva V, Kattla JJ, Flowers SA, Lindén SK, Premaratne P, Weijdegård B, Sundfeldt K,
1111 Karlsson NG 2015. The O-linked glycome and blood group antigens ABO on mucin-type

glycoproteins in mucinous and serous epithelial ovarian tumors. *PLoS ONE* 10:e0130197. DOI: 10.1371/journal.pone.0130197.

Vitti JJ, Grossman SR, Sabeti PC 2013. Detecting natural selection in genomic data. *Annual review of genetics* 47:97–120. DOI: 10.1146/annurev-genet-111212-133526.

Wang J, Chen L, Chen Z, Zhang W 2015. RNA-seq based transcriptomic analysis of single bacterial cells. *Integr. Biol.* 7:1466–1476. DOI: 10.1039/C5IB00191A.

Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui J-M, Bayne EH, Longdon B, Buck AH, Lazzaro BP, Akorli J, Haddrill PR, Obbard DJ 2015. The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. *PLoS Biology* 13:e1002210. DOI: 10.1371/journal.pbio.1002210.

Wegner KM, Reusch TBH, Kalbe M 2003. Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology* 16:224–232.

Westermann AJ, Barquist L, Vogel J 2017. Resolving host–pathogen interactions by dual RNA-seq. *PLoS Pathogens* 13:e1006033. DOI: 10.1371/journal.ppat.1006033.

Williams TN, Mwangi TW, Wambua S, Peto TEA, Weatherall DJ, Gupta S, Recker M, Penman BS, Uyoga S, Macharia A, Mwacharo JK, Snow RW, Marsh K 2005. Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait. *Nature Genetics* 37:1253–1257. DOI: 10.1038/ng1660.

Windsor DA 1998. Most of the species on Earth are parasites. *International journal for parasitology*. DOI: 10.1016/S0020-7519(98)00153-2.

Wohl S, Schaffner SF, Sabeti PC 2016. Genomic analysis of viral outbreaks. *Annual Review of Virology* 3:173–195. DOI: 10.1146/annurev-virology-110615-035747.

Wood CL, Johnson PT 2015. A world without parasites: exploring the hidden ecology of infection. *Frontiers in Ecology and the Environment* 13:425–434. DOI: 10.1890/140368.

Yuelong Shu JM 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* 22:957. DOI: 10.2807/1560-7917.ES.2017.22.13.30494.

Zhang Q, Hill GE, Edwards SV, Backström N 2014. A house finch (*Haemorhous mexicanus*) spleen transcriptome reveals intra- and interspecific patterns of gene expression, alternative splicing and genetic diversity in passerines. *BMC Genomics* 15:305. DOI: 10.1186/1471-2164-15-305.

Zhang Y, Lun C-Y, Tsui S 2015. Metagenomics: A new way to illustrate the crosstalk between infectious diseases and host microbiome. *International Journal of Molecular Sciences* 16:26263–26279. DOI: 10.3390/ijms161125957.

Table 1(on next page)

Table 1: Definition of categories for each scale and assigned scores used for the evaluation of host-pathogen literature.

Table 1: Definition of categories for each scale and assigned scores used for the evaluation of host-pathogen literature.

Score†	Genomic scale	Ecological scale	Temporal scale*	Spatial scale*
1	gene/ sequence fragment	none/ theoretical	none	none
2	full gene/ regulator	single species, laboratory system, environ. constant	single generation	local (one population)
3	gene family/ microsatellite	single species, laboratory system, environ. variable	few generations	intermediate (couple of populations)
4	whole plastid genome	multiple species, laboratory system, environment constant	many generations	species range
5	reduced genome representation	multiple species, laboratory system, environ. variable	speciation time (small tree)	global
6	exome/ transcriptome/ proteome	single species, natural system, environ. constant	speciation time (large tree)	
7	whole genome	single species, natural system, environ. variable		
8		multiple species, natural system, environ. constant		
9		multiple species, natural system, environ. variable		

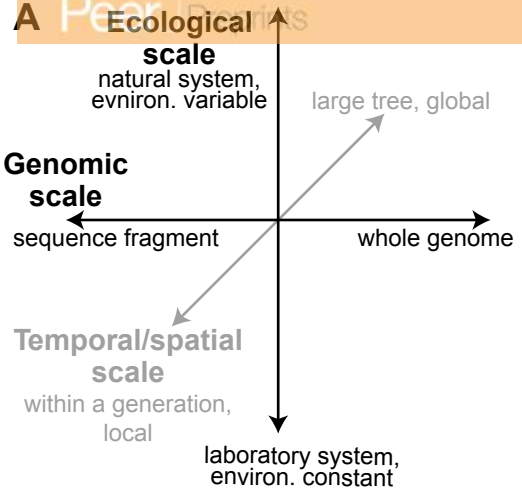
† see SI Table 1 for list of references and associated scoring results
* the spatiotemporal scale (Fig. 1) is the sum of the individual scores of the temporal and spatial scales

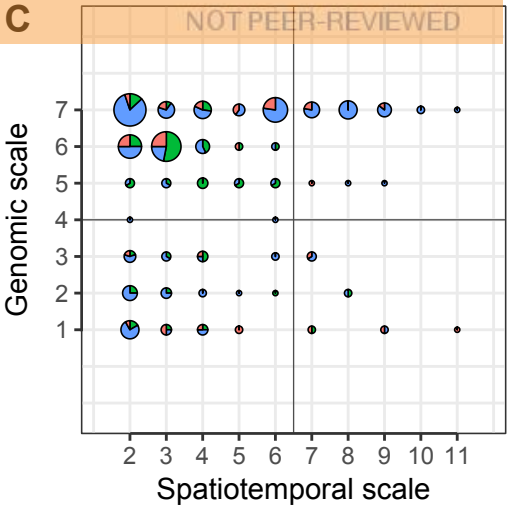
1

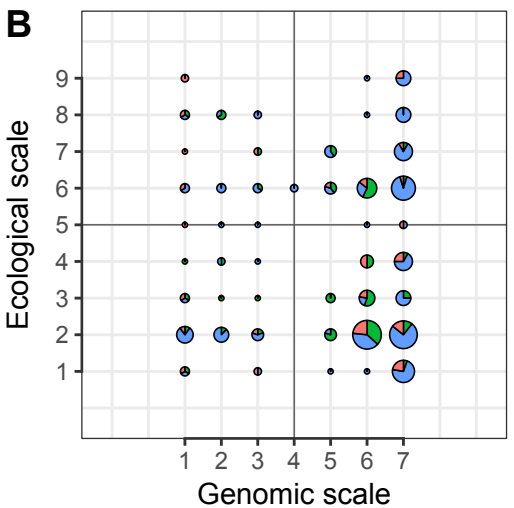
Figure 1(on next page)

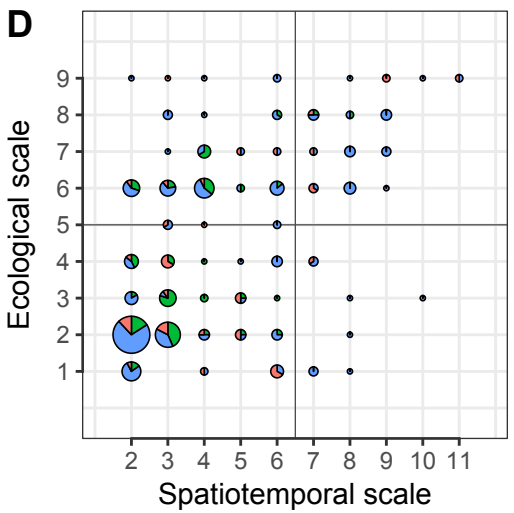
Figure 1: The diversity of recent studies of host-pathogen interactions.

(A) Each of three scales of complexity – genomic, ecological and spatiotemporal - is represented as an axis in this illustration. A study of host-pathogen interaction is placed into this three-dimensional space based on the level of genetic, ecological, and spatiotemporal detail that is being studied (see Table 1 for scores of scales). **(B-D)** Pie charts summarize the results of the scores for the level of genetic, ecological, and spatiotemporal complexity investigated in host-pathogen studies published between 2014-2018. **(B)** The complexity of the ecological and genomic settings across studies are not correlated (Spearman's $\rho = 0.02$, p-value adjusted = 1.00; **(C)** nor are the genomic and spatiotemporal scale ($\rho = 0.16$, p-value adj. = 0.13**. (D)** In contrast, the ecological scale positively correlates with the score of spatiotemporal scale across studies ($\rho = 0.50$, p-value adj. = 0.00).



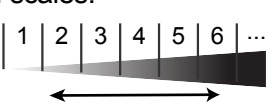






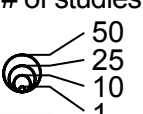
Study focus:
- Host; n = 61
- Pathogen; n = 159
- both; n = 43

Scores of scales:



lowest complexity (e.g. sequence fragment) — highest complexity (e.g. whole genome)

# of studies:

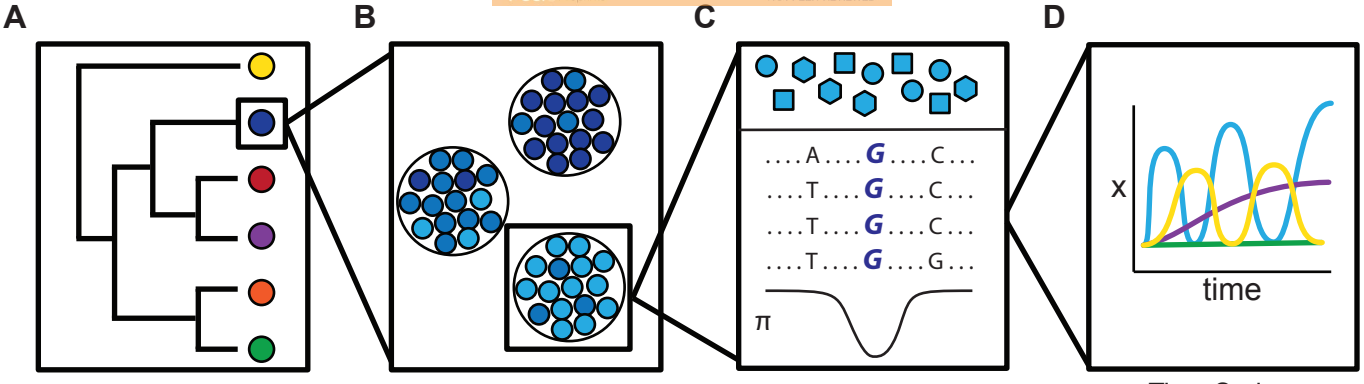


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27734v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

Figure 2(on next page)

Figure 2: Schematic illustration how genetic variation varies (A) across species, (B) across populations, (C) within a population, and (D) on an ecological time scale.

Comparative genomics across species can be used to identify genomic loci consistently under positive selection in particular lineages or all lineages **(A)**. Across populations **(B)**, population genomic variation in different geographic populations can be correlated with pathogen communities. Within a single population **(C)**, phenotypic variation among individuals can be linked to pathogen variation or differentially expressed genes with transcriptomics. Genome scans may also identify regions of the genome under selection. Finally, time series **(D)** either derived through experimental evolution or studies of ancient DNA or diachronic samples can be used to identify the dynamics of a phenotype or allele frequency through time.



Comparative Genomics

Population Genomics
Community Variation

Transcriptomics
Infection Experiments
Selection Scans

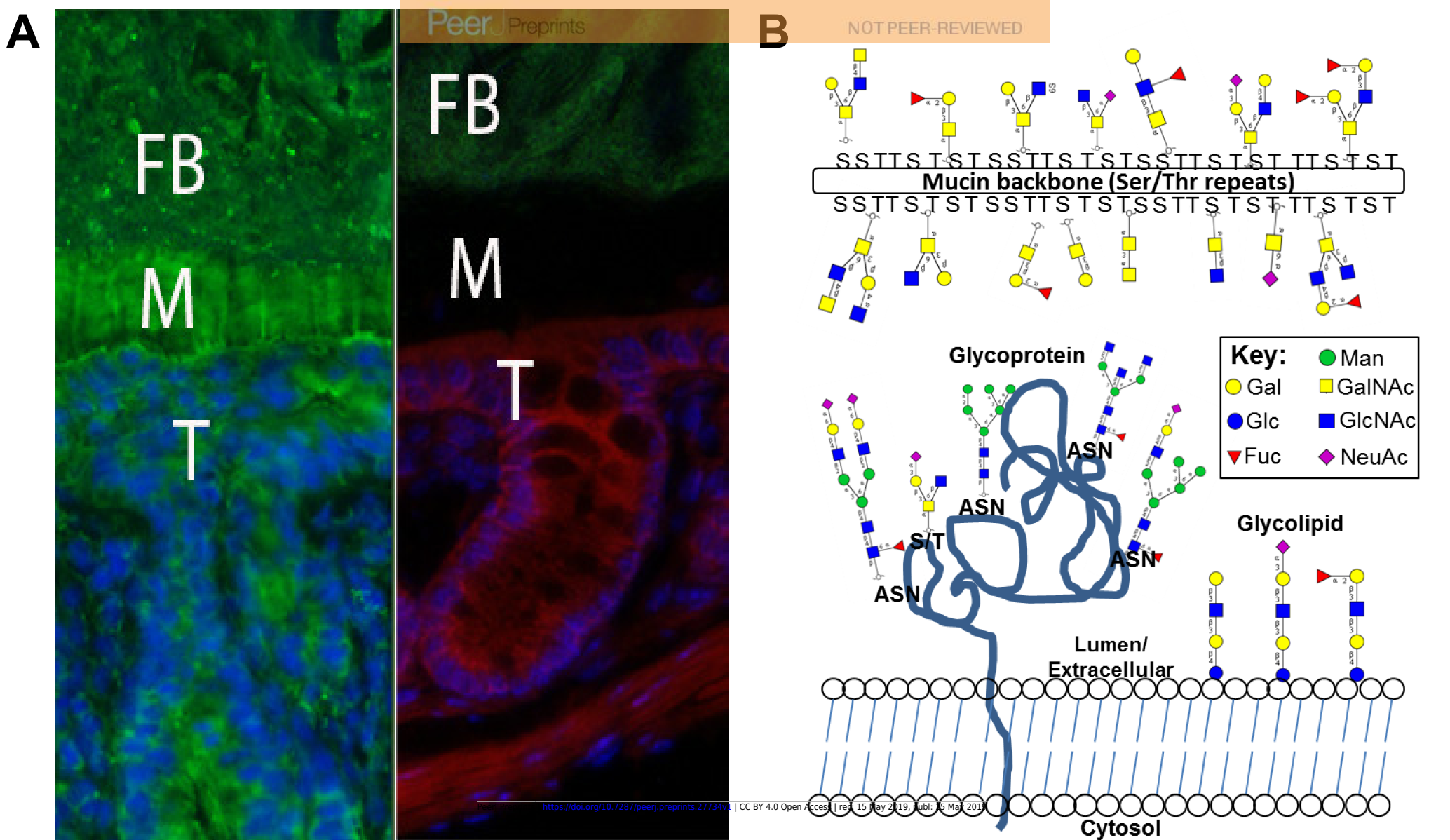
Time Series:
Experimental Evolution
Ancient DNA

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27734v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

Figure 3(on next page)

Figure 3: The mucosal layer.

(A) shows two staining variants of the colonic mucosal tissue (T) of a healthy mouse, where a mucus layer (M) keeps the majority of the fecal bacteria (FB) from direct contact with the surface of the epithelial cells. On the left side, the Muc2 mucin (the main component of the mucus layer) is stained in green and nuclei from the eukaryote cells in the tissue are stained blue. Muc2 is produced by cells in the mucosal tissue, secreted into the mucus layer, and present in degraded form in the fecal material. On the right side, the mucosal epithelial tissue is outlined with red, eukaryotic nuclei are purple, the mucus layer unstained (but clearly visible due to the absence of bacteria) and the bacteria are labelled green. Panel **(B)** gives an overview of glycan structures that build the mucus layer and glycocalyx. Glycolipids and glycoproteins are anchored in the eukaryotic cell membrane, and secreted mucins are highly glycosylated glycoproteins consisting of 70-90% of glycans that make up the bulk of the mucus layer. The glycans can be longer and more complex than depicted in this illustration. The glycans can be either *N*-linked (via Nitrogen in asparagine) or *O*-linked (via Oxygen in serine or threonine) to the protein core, and these two types of glycan chains differ with regards to biosynthetic pathway and structure.



https://doi.org/10.7287/peerj.preprints.27734v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019