*Parasitology*

**cambridge.org/par**

# Parasitological research in the molecular age

Christian Selbach[1,*], Fátima Jorge[1,*], Eddy Dowle[2,*], Jerusha Bennett[1], Xuhong Chai[1], Jean-François Doherty[1], Alan Eriksson[3], Antoine Filion[1], Eleanor Hay[1], Ryan Herbison[1], Jocelyn Lindner[1], Eunji Park[1], Bronwen Presswell[1], Brandon Ruehle[1], Paulo M. Sobrinho[4], Eleanor Wainwright[1] and Robert Poulin[1]

[1]Department of Zoology, University of Otago, 340 Great King St, Dunedin, 9016, New Zealand; [2]Department of Anatomy, University of Otago, 270 Great King St, Dunedin, 9016, New Zealand; [3]Programa de Pós-Graduação em Ecologia e Conservação, Universidade Federal de Mato Grosso do Sul, 79070-900 Campo Grande, Mato Grosso do Sul, Brazil and [4]Programa de Pós-Graduação em Etnobiologia e Conservação da Natureza, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, Dois Irmãos, 52171-900 Recife, PE, Brazil

**Review**

*Authors contributed equally to this work.

**Cite this article:** Selbach C *et al* (2019). Parasitological research in the molecular age. *Parasitology* **146**, 1361–1370. https://doi.org/10.1017/S0031182019000726

Received: 26 March 2019
Revised: 10 May 2019
Accepted: 16 May 2019
First published online: 18 June 2019

**Key words:**
Allozyme; genomics; molecular tools; nucleotide sequencing; publication trends

**Author for correspondence:**
Christian Selbach, E-mail: chris.selbach@gmail.com

## Abstract

New technological methods, such as rapidly developing molecular approaches, often provide new tools for scientific advances. However, these new tools are often not utilized equally across different research areas, possibly leading to disparities in progress between these areas. Here, we use empirical evidence from the scientific literature to test for potential discrepancies in the use of genetic tools to study parasitic *vs* non-parasitic organisms across three distinguishable molecular periods, the allozyme, nucleotide and genomics periods. Publications on parasites constitute only a fraction (<5%) of the total research output across all molecular periods and are dominated by medically relevant parasites (especially protists), particularly during the early phase of each period. Our analysis suggests an increasing complexity of topics and research questions being addressed with the development of more sophisticated molecular tools, with the research focus between the periods shifting from predominantly species discovery to broader theory-focused questions. We conclude that both new and older molecular methods offer powerful tools for research on parasites, including their diverse roles in ecosystems and their relevance as human pathogens. While older methods, such as barcoding approaches, will continue to feature in the molecular toolbox of parasitologists for years to come, we encourage parasitologists to be more responsive to new approaches that provide the tools to address broader questions.

## Introduction

Technological breakthroughs are often the product of scientific knowledge, and can in turn provide new tools for scientific advances. For instance, technology has empowered the study of life, from the invention of the microscope to the latest developments in genomics, imaging and GPS tracking technology. If biologists make rational and unconstrained decisions about what they study, we might expect that research effort would be allocated to different types of organisms in proportion to their diversity and abundance, their accessibility, their importance to society or to scientific progress, and how much funding they are likely to attract. However, much evidence exists to show that this is not always the case, and that allocation of research effort to different taxa is biased in many other ways (e.g. Hendriks and Duarte, 2008; Ahrends *et al*., 2011; Martin *et al*., 2012). This may also be reflected in the adoption of technological breakthroughs, with new research tools applied earlier and more frequently to certain taxa than others. In this synthesis, we examine discrepancies in the temporal deployment of new molecular methodologies toward the study of parasitic *vs* non-parasitic organisms, explore the possible reasons underlying the observed differences, and look ahead to the near future of molecular research on parasites.

A priori, there should be no reason why the application of molecular tools to parasitic organisms should have followed a different trajectory than their application to non-parasites. Most estimates agree that nearly half the living species are parasites, and that every free-living species (except perhaps some small unicellular taxa) harbours parasites (Windsor, 1998; Dobson *et al*., 2008; Poulin, 2014). Parasites are phylogenetically diverse, having evolved independently over 200 times among metazoan lineages alone (Weinstein and Kuris, 2016). They occur in all types of environment, surpassing their hosts in absolute abundance, and have been shown to play major roles in host population dynamics (Hudson *et al*., 1998; Tompkins *et al*., 2002; Møller, 2005), community structure (Mouritsen and Poulin, 2005; Wood *et al*., 2007; Hatcher and Dunn, 2011), food web stability and energy flow (Kuris *et al*., 2008; Lafferty *et al*., 2008; Selakovic *et al*., 2014; Preston *et al*., 2016). Parasites also have huge impacts on human health. Great white sharks, venomous snakes and killer bees make the headlines, but parasites kill orders of magnitude more people every year, and cause debilitating diseases in an even greater number of people (Torgerson *et al*., 2015). Parasites are also responsible for reduced production and huge economic losses in livestock farming (Rist *et al*., 2015), and pose

© Cambridge University Press 2019

CAMBRIDGE UNIVERSITY PRESS

serious challenges to conservation biologists and wildlife managers striving to protect the free-living species we cherish the most (Lafferty and Gerber, 2002; Smith *et al.*, 2009).

Parasites are therefore arguably as (if not more) diverse, ubiquitous and important as non-parasitic organisms. When new research tools become available, one would thus expect them to be applied to parasites no later than, and as frequently as, they are applied to non-parasites. For instance, the last several decades have witnessed the rapid expansion of molecular technologies as a means to explore the genetic basis of life. Molecular methods are now part of the standard toolkit of scientists in fields as diverse as evolutionary biology, taxonomy, ecology, developmental biology, immunology, medicine, and of course genetics. The growth of molecular technology has been regularly punctuated by major advances and their rapid conversion into usable tools. The molecular genetics era can be split into three distinct periods, each characterized by its own set of methods, analytical or statistical approaches and possibilities (Parker *et al.*, 1998; Schlötterer, 2004; Metzker, 2005; Schuster, 2007; Metzker, 2009). First, in the allozyme period, one could quantify differences in amino acid composition between enzymes from different individuals or species, as a proxy for genetic variation. Second, in the nucleotide sequencing period, it became possible and relatively simple to obtain actual DNA sequences for multiple genes, greatly expanding our window into an organism's genetic blueprint. Finally, in the genomics period, it is now possible to sequence entire genomes, as well as the transcriptomes, of living organisms, providing a complete picture of their genetic underpinnings.

These developments have provided researchers with an arsenal of new tools to explore the biology of parasites and non-parasites alike. Whereas there is no reason why these should be applied earlier or more frequently to either type of organisms, among parasites we may expect new molecular methods to be first used to study human parasites before they are applied to parasites of other organisms, simply because there is more funding and greater urgency to investigate the former in order to reduce human suffering. Furthermore, among both human parasites and those of other organisms, molecular tools may be applied disproportionately more to certain taxa than others, due to differences in diversity or perceived importance.

Here, we use empirical evidence from the scientific literature to test quantitatively the general hypothesis that the application of molecular tools to study parasites follows the same temporal profile as their application to the study of non-parasites. Also, among parasites, we test the hypothesis that new molecular tools are first adopted and used more extensively to study medically important species than any other type of parasites. More specifically, we answer the following questions: (i) Are new molecular tools adopted at the same time for the study of parasites and non-parasites, and if not, what is the time difference? (ii) How much of the early adoption of new molecular approaches is driven by research on medically important parasites? (iii) What are the most common topics of parasitological research using molecular approaches, and did these stay the same across the three periods of molecular research (allozyme, nucleotide, genomics)? (iv) What are the most commonly studied parasite taxa? and (v) What is the geographical distribution of molecular parasitological research? Our review provides a historical overview and illustration of the growth and development of molecular parasitology, as well as an exploration of the cultural differences between parasitologists and biologists studying non-parasites. We conclude with central take-home messages and recommendations for the future growth of parasitological research to assess the ecology and evolution of these phylogenetically diverse and ecologically important organisms.

## Materials and methods

### Data search and compilation

In order to analyse the research output on parasites in the molecular era, we identified and characterised three distinct periods of molecular research: the allozyme period, the nucleotide period and the genomics period. The 'allozyme period' is characterised by the use allozyme techniques (i.e. analysing differences in enzyme structure between organisms) that were established in the 1960's (Hubby and Lewontin, 1966; Schlötterer, 2004). First conceived in the 1970's–1980's but popularised in the 1990's, the 'nucleotide period' is largely characterised by the use of Sanger sequencing and microsatellites (Sanger *et al.*, 1977; Litt and Luty, 1989; Mathies and Huang, 1992; Richard *et al.*, 2008) and a small number of markers to distinguish between DNA sequence variations. Finally, the 'genomic period' is characterised by the onset of next generation sequencing (Solexa, 454, Illumina, SOLiD, etc.) established after 2005 (Bennett, 2004) and use of large datasets frequently utilising 1000s of markers on multiple genomic loci or whole genomes. Although treated as somewhat discrete periods in this study, in reality they are interwoven and overlap in time. For example, many genomes were initially sequenced using Sanger sequencing in the 1990s, but here whole genome research is classed as part of the genomic period.

For each period, we chose a set of keywords (see Supporting File S1) that captures the molecular markers, tools and methods developed and utilized in this context. We examined and compiled data on the publication trends of molecular research in these three periods by conducting a detailed search on the Web of Science™ for all entries until November 2018. In order to only capture relevant publications from the respective periods, we excluded the succeeding period(s) from the searches, e.g. when searching for the nucleotide period, we excluded search terms belonging to the genomics period. Moreover, the searches were restricted to relevant categories (i.e. biological, environmental, medical sciences, etc.), and to peer-reviewed research articles or reviews. For the analyses of publication trends in overall biological research (i.e. research on parasites and non-parasites together) within each period, we downloaded only the numbers of publications per year.

We then performed the same search for each molecular era, with the addition of a range of search terms for parasitic organisms to capture the majority of parasite-related publications within the three molecular periods. Although we did not include bacteria, fungi or viruses in these parasite-search terms, we did not exclude those groups specifically. For the analyses of publication trends on parasitic organisms, we downloaded the full records for all parasite-related publications from each molecular period (including title, abstract, keywords, author country, publication date and journal). The individual search terms for all molecular periods, the specific parasite-search terms and the Web of Science categories are presented in Supporting File S1.

### Analyses

All analyses were performed in R (R Core Team, 2018). Data from the searches was imported into R using the bibliometrix package (Aria and Cuccurullo, 2017). To determine what proportion of parasitology research is comprised of medically relevant parasites, we set up two categories of parasitological research, 'all parasites' and 'medical parasites'. All parasites contained the raw parasite data as downloaded from Web of Science. The second group, 'medical parasites', contained all the papers from within the all parasite group which were categorised as being medically focused. Assignment to the medically focused category was done by

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press

**Table 1.** Number of publications in the three periods of molecular research

| | *Allozyme* period | *Nucleotide* period | *Genomics* period | Total |
|---|---|---|---|---|
| General research | 19 418 | 1 046 366 | 264 699 | 1 330 483 |
| All parasites (% of general) | 823 (4.2) | 51 135 (4.9) | 9616 (3.6) | 61 574 (4.6) |
| Medically relevant parasites(% of all parasites) | 568 (69) | 34 697 (67.9) | 6281 (65.3) | 41 546 (67.5) |

searching across the papers' abstracts, titles and keywords for a list of terms associated with either humans (e.g. human, man, woman) or human-related pathogens and diseases (e.g. *Plasmodium falciparum*, *Plasmodium vivax*, lyme disease). Details of the terms are available in Supporting File S2. This resulted in three datasets across the three periods: 'general', 'all parasites' and 'medical parasites'.

To compare when and how fast new techniques were adopted across the three datasets, research publication trends within each era and group ('general, all parasites, medical parasites') were analysed from 1960–2017 by plotting the number of publications over time. To determine whether research focus had shifted across the three periods within parasitological research ('all parasites'), we generated word clouds and networks using publication keywords. Keywords were first cleaned to avoid duplications due to abbreviations (e.g. PCR and polymerase chain reaction) or pluralisation (e.g. tick and ticks). For the word clouds, we also removed all our search terms (e.g. 'nucleotide sequencing'), taxa or disease names (e.g. 'malaria' or '*Plasmodium*') and molecular markers (e.g. 'COI'). Word clouds were then generated using the top 40 keywords across each period in the wordcloud package in R (Fellows, 2018). Networks for each period were generated using the ggplot2 and ggraph packages in R (Wickham, 2016; Pedersen, 2018). Within each of the three periods, keyword pairs or words that occur together within an individual article's keywords were summed across all articles. Networks for each period were generated using the top 100 keyword pairs that had a minimum of three counts. To analyse which groups of parasites are most commonly studied using molecular tools, we quantified the proportion of papers published per taxonomic group within parasitological research ('all parasites') using a Sankey plot generated *via* plotly in R (Sievert, 2018). The focal taxon (or taxa) for each paper was determined by searching for taxa names across article description fields (title, abstract, keywords). Papers that mentioned more than one taxonomic group in the description fields were either reassigned to a single category (e.g. studies on arthropods as a vector of *Leishmania* were assigned to the latter), or classified as 'multiple species' (e.g. studies on both cestodes and trematodes were assigned to 'helminths' but not divided further). Finally, to visualize the geographic distribution of parasite research, world maps were generated for each molecular period with the R package rworldmap (South, 2011). For each publication the respective country of origin was defined according to the affiliation of the corresponding author.

The R code for all analyses is available in Supporting File S2.

## Results

The database search on the three periods of molecular research resulted in a total of 1.33 million publications, the majority of which appeared during the nucleotide period (Table 1). Studies on parasitic organisms constituted less than five percent of the total research output and were lowest in the genomics period.

The temporal publication trends show that parasite research in each molecular period followed the same overall patterns as the general output over the years (Fig. 1A). The sudden jump in publication numbers from 1990 to 1991 is an artefact of the Web of Science search algorithm (see Pautasso, 2014) and does not allow an accurate comparison of pre- and post-1990 publication patterns. However, since we are interested in the differences between overall publication trends and subsets of these datasets (studies on parasites and medically relevant parasites), and the artefact applies to all these groups equally, this does not distort our analyses. Post 1990, publication output in the nucleotide and genomics periods continually increased, while the allozyme dataset revealed a decreasing publication trend. Parasite research output followed these general trends but at a slightly more conservative rate, i.e. they show a slower increase in the nucleotide, and a slower decline in the allozyme periods (see Supporting Fig. S3). Within the parasite subsets across all periods, medically relevant parasites dominate the research output, making up 65–70% of publications. Across all periods of molecular research, early publications almost exclusively comprise studies on medically relevant parasites (Fig. 1B–D); the larger nucleotide and genomics datasets show that research on non-medical parasites only appears in reasonable numbers after a delay of several years, and starts to slowly increase thereafter.

Word clouds and networks of publication keywords show the predominant research fields and topics of parasitological research using the three molecular approaches (Figs 2 and 3; see Supporting File S4 for a large version of Fig. 3). Across the three periods, there is a shift in the main focus of research from discovery and taxonomy in the allozyme period to phylogeographic and disease diagnosis in the nucleotide period, to finally evolutionary genetics and comparative genomics in the genomic period. A large proportion of the top words in all periods are directly related to medical research, e.g. epidemiology, diagnosis. The most interconnected terms within the networks are related to the techniques that define each period (e.g. allozyme, polymerase chain reaction, genomics). In contrast with the later periods, medical research in the allozyme period focuses largely on *Leishmania* with no reference to *Plasmodium* and malaria.

The Sankey plot reveals the most commonly studied parasite taxa from all periods and illustrates the strong research focus on protist parasites, in particular *Plasmodium*, *Trypanosoma*, *Leishmania*, and *Toxoplasma* (Fig. 4, see Supporting File S5 for an interactive version of the figure). While research on protists remains dominant across all periods, there is a shift in its relative contribution from 56% in the allozyme period to 45 and 41% in the nucleotide and genomics period, respectively, while research interest in other parasite groups increases (e.g. arthropods, multiple species). The analysis of the geographical distribution of molecular parasitological research shows the vast majority of publications originate from the United States, with the exception of research during the allozyme period where Brazil and Europe were considerable contributors (Fig. 5, see Supporting File S6 for an interactive version of the figure). Although the United States remains the highest contributor, within the last 10 years, particularly the last 2 years, a growing number of publications originate from China within both the nucleotide and genomic periods.

## Discussion

Since parasites play central roles in all ecosystems and their evolutionary trajectory is closely linked with that of their hosts,

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press

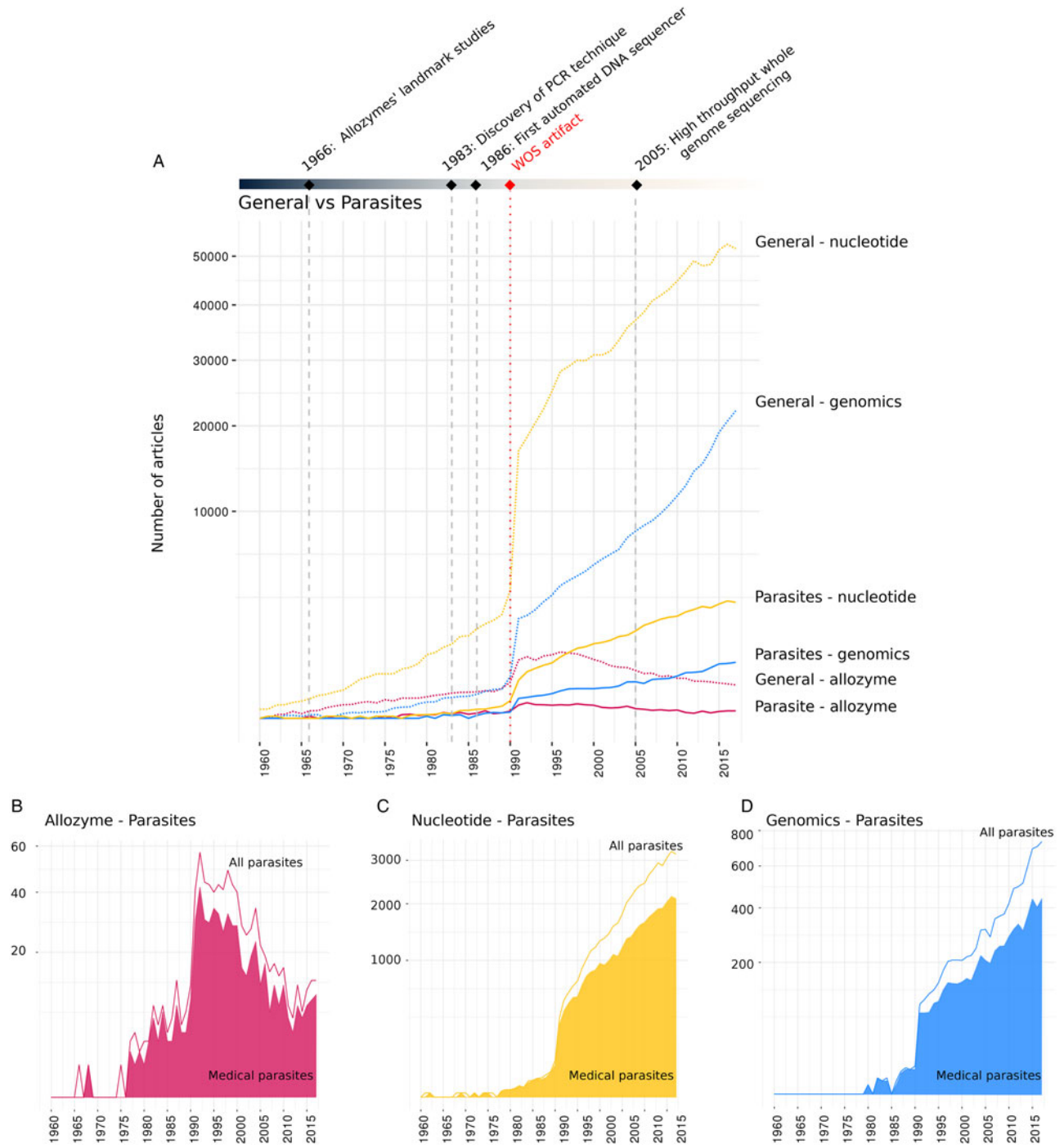



**Fig. 1.** Number of studies published during the three molecular periods from 1960 to 2017 (research articles and reviews). (A) General research output and parasite-related publications in each period; (B) Output of publications on parasites and medically relevant parasites in the allozyme period; (C) Output of publications on parasites and medically relevant parasites in the nucleotide period; (D) Output of publications on parasites and medically relevant parasites in the genomics period. The sudden jump in publication numbers from 1990 to 1991 is an artefact of the Web of Science search algorithm (see results).

research on parasitic organisms is central to our understanding of most fundamental biotic interactions and concepts. Parasitic life styles have evolved multiple times, and parasitic organisms are estimated to account for 30–50% of global biodiversity (Windsor, 1998; Dobson *et al.*, 2008; Poulin, 2014; Weinstein and Kuris, 2016). We therefore hypothesized that new molecular tools would be applied to parasites no later than, and as frequently as, they are applied to non-parasites.

Our results suggest a more complex picture. Only less than 5% of the 1.3 million publications in our dataset deal with parasitic organisms. Moreover, within these publications on parasites, almost 70% of the studies focused on medically relevant taxa. In particular, protists of the genera *Plasmodium*, *Trypanosoma*, *Leishmania*, and *Toxoplasma*, as well as mites and ticks (Acari) and the disease agents transmitted by these vectors, have been the predominantly studied parasite groups. Other important disease agents attract far less attention. In fact, the molecular research output on *Plasmodium* alone (10 800 publications) exceeds the combined publication volume on all helminths (10 600 publications). Likewise, not all so-called neglected tropical diseases (NTDs) receive similar research attention; while *Trypanosoma* (3750 publications), *Leishmania* (2930 publications) and *Schistosoma* (2100 publications) are relatively well-studied groups, other pathogens, such as *Ascaris* and hookworms

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press

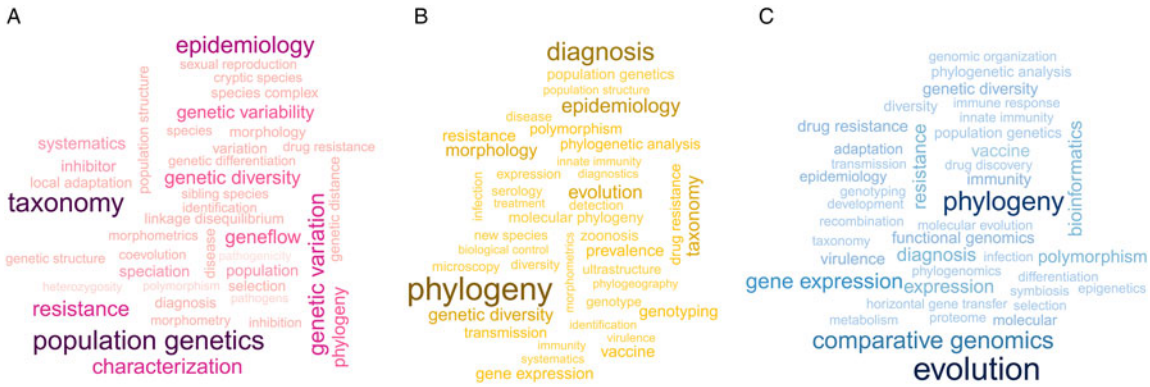



**Fig. 2.** Word clouds highlighting the most common research fields in the individual molecular periods. (A) Allozyme period; (B) nucleotide period; (C) genomics period. The more frequently a term is used in the keywords, the larger and darker it is shown in the word cloud.

receive far less attention (550 and 200 publications, respectively), despite more than a quarter of the world's population being at risk of infection with these soil-transmitted helminths (Jourdan *et al.*, 2018). Overall, these publication trends in molecular research on parasites seem to support recent findings that research effort on different parasitic pathogens does not reflect the actual human disease burden they inflict, and highlights that some medically relevant taxa remain severely understudied (Furuse, 2019).

Beyond this uneven research focus within the group of medically important parasites, our results show that the vast majority of parasite species are not studied with these molecular tools, regardless of whether the true diversity of parasitic organisms is closer to the higher or lower end of current diversity predictions (see Poulin, 2014). Due to the large dataset with over one million published articles and similar relative numbers of medically relevant parasites across all periods of molecular research, we believe our results are representative of the general trends in parasite research. Nevertheless, it would be interesting to see if this discrepancy between predicted biodiversity of parasites and research output, as well as the strong focus on medically important taxa, are similarly pronounced for other methodological approaches, such as electron microscopy and modern imaging technologies.

The comparison of general and parasite-related publication output in the scientific literature indicates some dissimilarities in the temporal trends, i.e. in the nucleotide period, parasite research output showed a slower increase relative to the general field, whereas the output in the allozyme period decreased at a slower rate than the general field. This might suggest that parasite research is slightly more conservative in using these molecular tools. It could be that the way in which parasitologists were trained led to possible cultural differences between them and biologists studying non-parasites, influencing their willingness to adopt new techniques from other fields. However, it must be noted that no difference was apparent in the rapidly developing genomics field.

Given the substantial global burden parasitic diseases impose on humanity (see GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016), it is not surprising that new molecular methods are first used to study these medically relevant parasites, in order to better understand their biology and distribution, and develop control approaches, treatments and vaccines to reduce human suffering. While during the 1990s, cost per megabase (mb, i.e. a million bases) exceeded 10 000 USD and initial genome projects ran into the hundreds of millions (the human genome project alone is estimated at 2.7 billion USD), recent years have seen the cost of sequencing drastically reduced (2017 estimate <0.1 USD per mb) (Hall, 2013; National Human Genome Research Institute, 2018). Moreover, beyond sequencing costs, the increasingly complex datasets in the genomics period require thorough bioinformatic analyses that needed to become more accessible (Muir *et al.*, 2016). This would explain why, in both the nucleotide and genomics periods, it took years for research on parasites other than medically important ones to get started.

We were interested in finding out what topics parasitological research addresses using molecular approaches, and whether these stayed the same across the three periods of molecular research. The word clouds reveal a gradual shift from predominantly species discovery (taxonomy and population genetics) during the allozyme period to studying the relationships among various species (phylogeny) during the nucleotide period, and finally addressing broader theory-focused questions (evolution) in the genomics period. Correspondingly, the networks suggest an increasing complexity of topics and research questions addressed with more sophisticated methods from the allozyme to the genomics period. Additionally, the networks reveal more detailed changes in the application and research foci of the different molecular methods. While medical research in the allozyme period focuses largely on *Trypanosoma* and *Leishmania* with no reference to malaria, the later nucleotide sequencing and genomics periods show a greater diversity of disease-related studies, most likely due to the larger number of studies in the datasets, and a strong emphasis on *Plasmodium* and malaria research.

The distinctive cluster in the nucleotide era network (Fig. 3B) that links the keywords 'morphology', 'taxonomy' and 'phylogeny' highlights the common use of integrative molecular and morphological approaches to identify and describe new parasite species. Indeed, 'integrative taxonomy' has replaced classical taxonomy in the early 2000s, right in the middle of the nucleotide period (Dayrat, 2005). Nucleotide sequencing approaches (e.g. barcoding) have become practical, reliable and cost-efficient tools to address these issues and uncover hidden and cryptic parasite diversity in ecosystems (e.g. Locke *et al.*, 2010; Selbach *et al.*, 2015). Moreover, many parasites have complex life cycles with morphologically highly distinct stages in different hosts; these stages are therefore difficult to match and assign to the same species (Poulin and Leung, 2010; Locke *et al.*, 2013). Consequently, many parasite life cycles are not yet fully described, and modern molecular tools are proving invaluable to resolve life cycles (e.g. Jensen and Bullard, 2010; Blasco-Costa *et al.*, 2016). Since much of parasite diversity is still unknown and our discovery efforts lag well behind our knowledge of free-living diversity (see Jorge and Poulin, 2018), these well-established molecular approaches are helping to close this gap.

Despite the global disease burden caused by many parasite taxa, the geographical distribution of molecular parasitological

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press

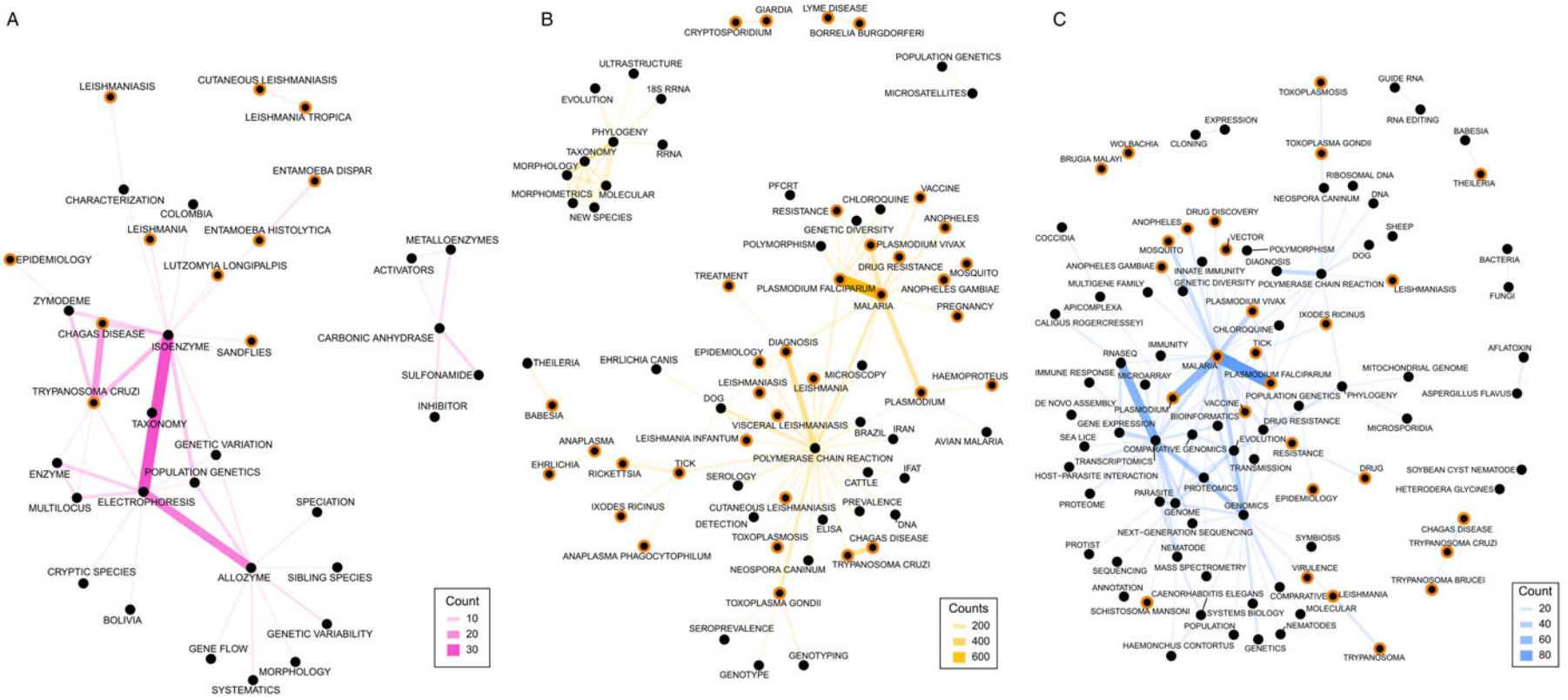



**Fig. 3.** Semantic networks for the individual molecular periods. (A) Allozyme period; (B) nucleotide period; (C) genomics period. Networks are based on terms used in the keywords. The more frequently terms co-occur within the same paper, the bolder and darker the connecting lines (see individual legends). Nodes that represent medically relevant terms are highlighted in orange.

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press



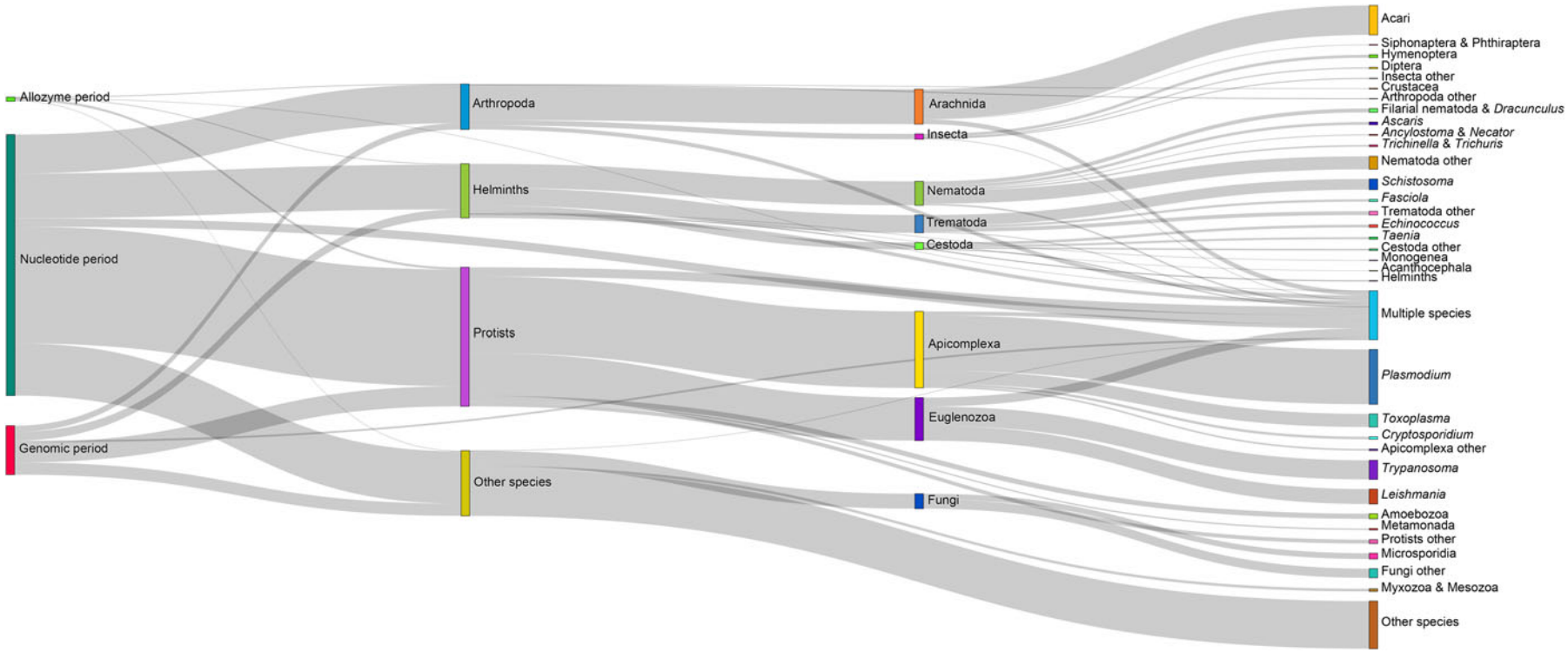



**Fig. 4.** Sankey plot showing the parasite taxa studied using molecular methods. The thickness of each line represents the relative number of publications on that particular taxon.

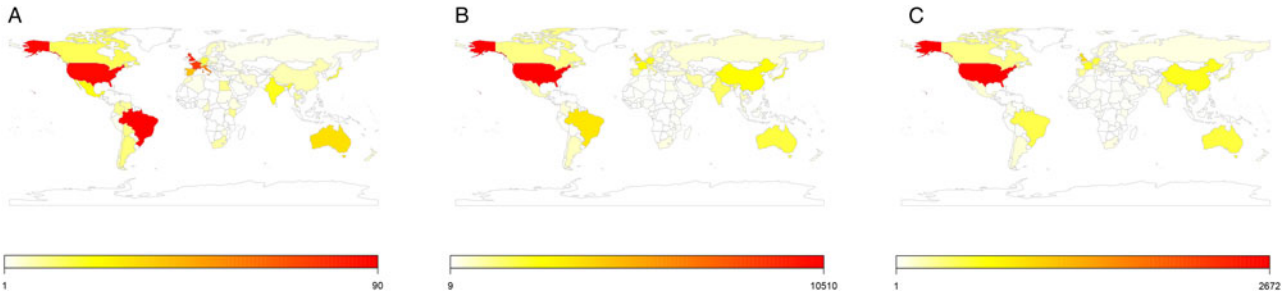



**Fig. 5.** Maps showing the publication output (research articles and reviews) during the three molecular periods from 1960 to 2017. (A) Allozyme period; (B) nucleotide period; (C) genomics period. Based on the affiliation of the corresponding author.

research is highly uneven. Across all three periods, the research landscape is dominated by a few individual countries, in particular the United States. Only the allozyme period shows a wider distribution with Brazil and Europe (France and the United Kingdom) contributing markedly to the overall research output. The higher number of studies from Brazil can likely be explained by the strong research focus of allozyme studies on *Leishmania* spp. and *Trypanosoma cruzi* that cause leishmaniasis and Chagas disease, respectively. Both represent important public health problems in Brazil, with Chagas disease alone resulting in an average of 5000 deaths per year (Alvar *et al.*, 2012; Ferro e Silva *et al.*, 2018). It must be noted however that our assessment of the geographical distribution of research is based on the affiliation of the corresponding author and therefore typically shows the origin of the funding rather than the study region or the location of collaborators (e.g. a study on malaria vectors in Venezuela by a senior author based at a research institute in the United States will be associated with the latter). Although scientific instruments, such as thermocyclers and sequencers, have become more readily available and cheaper and the overall costs for molecular analyses have decreased, in less economically developed regions such equipment and consumables often remain highly expensive (see van Helden, 2012), making it even harder for researchers in these countries to utilize molecular tools.

With countries in other regions, especially China in East Asia, increasing their research and development expenditure (UIS, 2019; World Bank, 2019) and many of the world's largest genomic institutes and projects, e.g. Beijing Genomics Institute and Genome Asia 100k, being focused on Asia, we expect a further shift in these patterns in the future. Looking at the research output from the last five years supports this and reveals a strong increase in publications from China in the nucleotide and genomics fields. Moreover, since some major parasitic disease agents, such as *Schistosoma japonicum* (Gryseels *et al.*, 2006; Wang *et al.*, 2009) or *Toxoplasma gondii* (Zhou *et al.*, 2011) are of public health and economic importance in Asia, we expect an increased research focus on these parasites and their vectors.

We predict that genomic research on parasites will continue to expand as part of the search for new anti-parasite drugs and vaccines, but parasitological research using Sanger sequencing will not decline as fast as it might for non-parasitic organisms. In contrast to many groups of free-living organisms, much of parasite biodiversity remains to be found and recorded. Accordingly, widely used gene markers (such as the mitochondrial COI or the nuclear 28S rRNA genes) will continue to play key roles in parasite discovery and taxonomy for years to come.

At the same time, the slower adoption of new ideas or technologies in basic parasitological research than in general biology should be remedied. New molecular approaches, such as environmental or eDNA analysis, where DNA (or RNA) is extracted from environmental or organismal matrices, provide promising tools for the study of parasitic organisms (Bass *et al.*, 2015). However, although eDNA surveys have found diverse applications to study free-living organisms (e.g. Thomsen and Willerslev, 2015; Stat *et al.*, 2017; Hering *et al.*, 2018; Lacoursière-Roussel *et al.*, 2018), parasitologists have so far been slow to adopt these methods, with a few exceptions (e.g. Huver *et al.*, 2015; Carraro *et al.*, 2018; Rusch *et al.*, 2018). This slow adoption is apparent in other areas, too. For example, although the microbiomes of parasites are now recognised as hugely important for understanding their biology and controlling them (Jenkins *et al.*, 2019), a coordinated effort to characterise and analyse parasite microbiomes is only a recent initiative (Dheilly *et al.*, 2017), appearing years after the establishment of other large-scale, organised programmes such as the Human Microbiome Project, or the Earth Microbiome Project.

One way for the faster incorporation of new developments in the study of parasites would be for the walls isolating different research areas to come down. There exists no society, conference or journal devoted exclusively to the study of non-parasitic organisms. Indeed, conferences and journals of ecology, evolutionary biology, or molecular genetics welcome papers on both parasites and non-parasites. In contrast, there are multiple societies, conferences and journals of parasitology, focused solely on the study of parasites. This one-sided compartmentalisation of the study of life may be responsible for the delayed integration of technologies developed originally for the study of non-parasites, by researchers focused on the study of parasites. The onus is therefore on parasitologists to reach across and fully join the wider biological research community.

Progress will come from the right blend of old and new technologies. However, novel approaches and methods need to infiltrate parasitological research faster than they have for the field to meet its future challenges, from making headway with the discovery of parasite biodiversity to the mitigation of their impact on human health.

## Conclusions

(i) Our analysis of publication trends shows that parasite-focused research in the three molecular periods (allozyme, nucleotide, genomics) follows the same overall patterns as the general biological research over the years but at a slightly more conservative rate. Despite the great diversity as well as ecological and medical importance of parasites, the total number of studies on parasitic organisms constitutes less than five percent of the total research output across all molecular periods.

(ii) Medically relevant parasites, in particular protists, dominate parasitology research, making up almost 70% of publications. Across all periods of molecular research, early publications almost exclusively comprise studies on medically relevant

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press

parasites and research on non-human parasites only appears in significant numbers after a delay of several years.

(iii) Our analysis reveals a gradual shift in research focus between the three periods, from largely species discovery studies (taxonomy and population genetics) in the allozyme period, to investigating the relationships among various species (phylogeny) during the nucleotide period, and finally addressing broader theory-focused questions (evolution) in the genomics period. Altogether, this suggests an increasing complexity of topics and research questions that can be addressed with the development of more sophisticated molecular tools.

(iv) With the exception of the allozyme period, the research output on molecular parasite research is dominated by authors affiliated with, and presumably financed by, institutions in the United States. Molecular tools are now far more cost-effective and accessible to researchers around the world, and the geographic distribution has begun to shift in recent years as other regions, particularly China, develop their genomic research.

(v) Altogether, we conclude that molecular methods provide powerful tools for research on parasitic organisms, including their diverse roles in ecosystems and their importance as human pathogens. Older methods, such as barcoding approaches using the COI gene, will continue to provide valuable items in the molecular toolbox for parasite research for years to come, since much of parasite biodiversity is still undiscovered. At the same time, we encourage researchers to be on the lookout for, and quickly integrate, novel approaches and methods to advance research on parasitic organisms.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/S0031182019000726.

**Author ORCIDs.** Christian Selbach, 0000-0002-7777-6515; Fátima Jorge, 0000-0002-3138-1729; Alan Eriksson, 0000-0003-1857-7935; Ryan Herbison, 0000-0001-5335-2382; Robert Poulin, 0000-0003-1390-1206.

**Acknowledgements.** We are grateful for the support from Emerson's Brewing Company. We also thank two anonymous referees for their comments and suggestions that improved the manuscript.

**Financial support.** This work was supported by a research fellowship from the German Research Foundation DFG (C.S., Grant Number: SE 2728/1-1, and SE 2728/2-1), a Marsden Fund awarded to R.P. (F.J., B.P.), University of Otago Doctoral Scholarships (J.F.D., A.F., E.P., B.R.), a University of Otago Summer Research Scholarship (X.C.), a University of Otago Masters Scholarship (R.H.), and financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES) (P.M.S., A.E., Grant Number: 88881.187634/2018-01).

**Conflict of interest.** None.

**Ethical standards.** Not applicable.

## References

**Ahrends A, Burgess ND, Gereau RE, Marchant R, Bulling MT, Lovett JC, Platts PJ, Kindemba VW, Owen N, Fanning E and Rahbek C** (2011) Funding begets biodiversity. *Diversity and Distributions* **17**, 191–200.

**Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J and den Boer M** (2012) Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* **7**, e35671.

**Aria M and Cuccurullo C** (2017) Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics* **11**, 959–975.

**Bass D, Stentiford GD, Littlewood DTJ and Hartikainen H** (2015) Diverse applications of environmental DNA methods in parasitology. *Trends in Parasitology* **31**, 499–513.

**Bennett S** (2004) Solexa Ltd. *Pharmacogenomics* **5**, 433–438.

**Blasco-Costa I, Poulin R and Presswell B** (2016) Species of *Apatemon* szidat, 1928 and *Australapatemon* sudarikov, 1959 (Trematoda: Strigeidae) from New Zealand: linking and characterising life cycle stages with morphology and molecules. *Parasitology Research* **115**, 271–289.

**Carraro L, Hartikainen H, Jokela J, Bertuzzo E and Rinaldo A** (2018) Estimating species distribution and abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences* **115**, 11724–11729.

**Dayrat B** (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85**, 407–415.

**Dheilly NM, Bolnick D, Bordenstein S, Brindley PJ, Figuères C, Holmes EC, Martinez JM, Phillips AJ, Poulin R and Rosario K** (2017) Parasite Microbiome Project: systematic investigation of microbiome dynamics within and across parasite-host interactions. *mSystems* **2**, e00050–17.

**Dobson A, Lafferty KD, Kuris AM, Hechinger RF and Jetz W** (2008) Homage to linnaeus: how many parasites? how many hosts? *Proceedings of the National Academy of Sciences* **105**, 11482–11489.

**Fellows I** (2018) wordcloud: Word Clouds. R package version 2.6. Available at https://CRAN.R-project.org/package=wordcloud.

**Ferro e Silva AM, Sobral-Souza T, Vancine MH, Muylaert RL, de Abreu AP, Pelloso SM, de Barros Carvalho MD, de Andrade L, Ribeiro MC and de Toledo MJO** (2018) Spatial prediction of risk areas for vector transmission of *Trypanosoma cruzi* in the State of Paraná, southern Brazil. *PLoS Neglected Tropical Diseases* **12**, e0006907.

**Furuse Y** (2019) Analysis of research intensity on infectious disease by disease burden reveals which infectious diseases are neglected by researchers. *Proceedings of the National Academy of Sciences* **116**, 478–483.

**GBD 2015 Disease and Injury Incidence and Prevalence Collaborators** (2016) Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* **388**, 1545–1602.

**Gryseels B, Polman K, Clerinx J and Kestens L** (2006) Human schistosomiasis. *The Lancet* **368**, 1106–1118.

**Hall N** (2013) After the gold rush. *Genome Biology* **14**, 115.

**Hatcher MJ and Dunn AM** (2011) *Parasites in Ecological Communities: From Interactions to Ecosystems*. Cambridge, UK: Cambridge University Press.

**Hendriks IE and Duarte CM** (2008) Allocation of effort and imbalances in biodiversity research. *Journal of Experimental Marine Biology and Ecology* **360**, 15–20.

**Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A and Kelly M** (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research* **138**, 192–205.

**Hubby JL and Lewontin RC** (1966) A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 577–594.

**Hudson PJ, Dobson AP and Newborn D** (1998) Prevention of population cycles by parasite removal. *Science* **282**, 2256–2258.

**Huver JR, Koprivnikar J, Johnson PTJ and Whyard S** (2015) Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic ecosystems. *Ecological Applications* **25**, 991–1002.

**Jenkins TP, Brindley PJ, Gasser RB and Cantacessi C** (2019) Helminth microbiomes: a hidden treasure trove? *Trends in Parasitology* **35**, 13–22.

**Jensen K and Bullard SA** (2010) Characterization of a diversity of tetraphyllidean and rhinebothriidean cestode larval types, with comments on host associations and life-cycles. *International Journal for Parasitology* **40**, 889–910.

**Jorge F and Poulin R** (2018) Poor geographical match between the distributions of host diversity and parasite discovery effort. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20180072.

**Jourdan PM, Lamberton PHL, Fenwick A and Addiss DG** (2018) Soil-transmitted helminth infections. *The Lancet* **391**, 252–265.

**Kuris AM, Hechinger RF, Shaw JC, Whitney KL, Aguirre-Macedo L, Boch CA, Dobson AP, Dunham EJ, Fredensborg BL, Huspeni TC, Lorda J, Mabada L, Mancini FT, Mora AB, Pickering M, Talhouk NL, Torchin ME and Lafferty KD** (2008) Ecosystem energetic implications of parasite and free-living biomass in three estuaries. *Nature* **454**, 515–518.

**Lacoursière-Roussel A, Howland K, Normandeau E, Grey EK, Archambault P, Deiner K, Lodge DM, Hernandez C, Leduc N and Bernatchez L** (2018) eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *International Journal of Business Innovation and Research* **17**, 7763–7777.

**Lafferty KD and Gerber LR** (2002) Good medicine for conservation biology: the intersection of epidemiology and conservation biology. *Conservation Biology* **16**, 593–604.

**Lafferty KD, Allesina S, Arim M, Briggs CJ, De Leo G, Dobson AP, Dunne JA, Johnson PTJ, Kuris AM, Marcogliese DJ, Martinez ND, Memmott J, Marquet PA, McLaughlin JP, Mordecai EA, Pascual M, Poulin R and Thieltges DW** (2008) Parasites in food webs: the ultimate missing links. *Ecology Letters* **11**, 533–546.

**Litt M and Luty JA** (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**, 397–401.

**Locke SA, McLaughlin JD, Dayanandan S and Marcogliese D** (2010) Diversity and specificity in *Diplostomum* spp. metacercariae in freshwater fishes revealed by cytochrome c oxidase I and internal transcribed spacer sequences. *International Journal for Parasitology* **40**, 333–343.

**Locke SA, McLaughlin JD and Marcogliese DJ** (2013) Predicting the similarity of parasite communities in freshwater fishes using the phylogeny, ecology and proximity of hosts. *Oikos* **122**, 73–83.

**Martin LJ, Blossey B and Ellis E** (2012) Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment* **10**, 195–201.

**Mathies RA and Huang XC** (1992) Capillary array electrophoresis: an approach to high-speed, high-throughput DNA sequencing. *Nature* **359**, 167–169.

**Metzker ML** (2005) Emerging technologies in DNA sequencing. *Genome Research* **15**, 1767–1776.

**Metzker ML** (2009) Sequencing technologies — the next generation. *Nature Reviews Genetics* **11**, 31–46.

**Møller AP** (2005) Parasitism and the regulation of host populations. In Thomas F, Renaud F and Guégan J-F (eds), *Parasitism and Ecosystems*. Oxford, UK: Oxford University Press, pp. 43–53.

**Mouritsen KN and Poulin R** (2005) Parasites boost biodiversity and change animal community structure by trait-mediated indirect effects. *Oikos* **108**, 344–350.

**Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, Zhang J, Weinstock GM, Isaacs F, Rozowsky J and Gerstein M** (2016) The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biology* **17**, 53.

**National Human Genome Research Institute** (2018) DNA Sequencing Costs: Data. Available at https://www.genome.gov/27541954/dna-sequencing-costs-data/.

**Parker PG, Snow AA, Schug MD, Booton GC and Fuerst PA** (1998) What molecules can tell us about populations: choosing and using a molecular marker. *Ecology* **79**, 361–382.

**Pautasso M** (2014) The jump in network ecology research between 1990 and 1991 is a Web of Science artefact. *Ecological Modelling* **286**, 11–12.

**Pedersen TL** (2018) ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. R package version 1.0.2. Available at https://CRAN.R-project.org/package=ggraph.

**Poulin R** (2014) Parasite biodiversity revisited: frontiers and constraints. *International Journal for Parasitology* **44**, 581–589.

**Poulin R and Leung TLF** (2010) Taxonomic resolution in parasite community studies: are things getting worse? *Parasitology* **137**, 1967–1973.

**Preston DL, Mischler JA, Townsend AR and Johnson PTJ** (2016) Disease ecology meets ecosystem science. *Ecosystems* **19**, 737–748.

**R Core Team** (2018) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R foundation for Statistical Computing. Available at http//www.R-project.org.

**Richard G-F, Kerrest A and Dujon B** (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews* **72**, 686–727.

**Rist CL, Garchitorena A, Ngonghala CN, Gillespie TR and Bonds MH** (2015) The burden of livestock parasites on the poor. *Trends in Parasitology* **31**, 527–530.

**Rusch JC, Hansen H, Strand DA, Markussen T, Hytterød S and Vrålstad T** (2018) Catching the fish with the worm: a case study on eDNA detection of the monogenean parasite *Gyrodactylus salaris* and two of its hosts, Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). *Parasites and Vectors* **11**, 1–12.

**Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison CA, Slocombe PM and Smith M** (1977) Nucleotide sequence of bacteriophage φX174 DNA. *Nature* **265**, 687–695.

**Schlötterer C** (2004) The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics* **5**, 63–69.

**Schuster SC** (2007) Next-generation sequencing transforms today's biology. *Nature Methods* **5**, 16–18.

**Selakovic S, de Ruiter PC and Heesterbeek H** (2014) Infectious disease agents mediate interaction in food webs and ecosystems. *Proceedings of the Royal Society B* **281**, 20132709.

**Selbach C, Soldánová M, Georgieva S, Kostadinova A and Sures B** (2015) Integrative taxonomic approach to the cryptic diversity of *Diplostomum* spp. in lymnaeid snails from Europe with a focus on the '*Diplostomum mergi*' species complex. *Parasites & Vectors* **8**, 300.

**Sievert C** (2018) plotly for R. Available at https://plotly-book.cpsievert.me.

**Smith KF, Acevedo-Whitehouse K and Pedersen AB** (2009) The role of infectious diseases in biological conservation. *Animal Conservation* **12**, 1–12.

**South A** (2011) Rworldmap: a New R package for Mapping Global Data. *The R Journal* **3**, 35–43.

**Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, Harvey ES and Bunce M** (2017) Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports* **7**, 12240.

**Thomsen PF and Willerslev E** (2015) Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**, 4–18.

**Tompkins DM, Dobson AP, Arneberg P, Begon ME, Cattadori IM, Greenman JV, Heesterbeek JAP, Hudson PJ, Newborn D, Pugliese A, Rizzoli AP, Rosà R, Rosso F and Wilson K** (2002) Parasites and host population dynamics. In Hudson PJ, Rizzoli A, Grenfell BT, Heesterbeek H and Dobson AP (eds), *The Ecology of Wildlife Diseases*. Oxford, UK: Oxford University Press, pp. 45–62.

**Torgerson PR, Devleesschauwer B, Praet N, Speybroeck N, Willingham AL, Kasuga F, Rokni MB, Zhou X-N, Fèvre EM, Sripa B, Gargouri N, Fürst T, Budke CM, Carabin H, Kirk MD, Angulo FJ, Havelaar A and de Silva N** (2015) World Health Organization estimates of the global and regional disease burden of 11 foodborne parasitic diseases, 2010: a data synthesis. *PLoS Medicine* **12**, e1001920.

**UNESCO Institute for Statistics (UIS)** (2019) How much does your country invest in R&D? Available at http://uis.unesco.org/apps/visualisations/research-and-development-spending/.

**Van Helden P** (2012) The cost of research in developing countries. *EMBO Reports* **13**, 395.

**Wang L-D, Guo J-G, Wu X-H, Chen H-G, Wang T-P, Zhu S-P, Zhang Z-H, Steinmann P, Yang G-J, Wang S-P, Wu Z-D, Wang L-Y, Hao Y, Bergquist R, Utzinger J and Zhou XN** (2009) China's new strategy to block *Schistosoma japonicum* transmission: experiences and impact beyond schistosomiasis. *Tropical Medicine & International Health* **14**, 1475–1483.

**Weinstein SB and Kuris AM** (2016) Independent origins of parasitism in Animalia. *Biology Letters* **12**, 20160324.

**Wickham H** (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

**Windsor DA** (1998) Most of the species on Earth are parasites. *International Journal for Parasitology* **28**, 1939–1941.

**Wood CL, Byers JE, Cottingham KL, Altman I, Donahue MJ and Blakeslee AMH** (2007) Parasites alter community structure. *Proceedings of the National Academy of Sciences* **104**, 9335–9339.

**World Bank** (2019) Research and development expenditure. Available at https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS.

**Zhou P, Chen Z, Li H-L, Zheng H, He S, Lin R-Q and Zhu X-Q** (2011) *Toxoplasma gondii* infection in humans in China. *Parasites & Vectors* **4**, 165.

https://doi.org/10.1017/S0031182019000726 Published online by Cambridge University Press