Origin of a major infectious disease in vertebrates: The timing of *Cryptosporidium* evolution and its hosts

JUAN C. GARCIA-R* *and* DAVID T. S. HAYMAN

Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Private Bag, 11 222, Palmerston North 4442, New Zealand

(*Received 10 April 2016; revised 14 June 2016; accepted 1 July 2016; first published online 30 August 2016*)

SUMMARY

Protozoan parasites of the genus *Cryptosporidium* infect all vertebrate groups and display some host specificity in their infections. It is therefore possible to assume that *Cryptosporidium* parasites evolved intimately aside with vertebrate lineages. Here we propose a scenario of *Cryptosporidium*–Vertebrata coevolution testing the hypothesis that the origin of *Cryptosporidium* parasites follows that of the origin of modern vertebrates. We use calibrated molecular clocks and cophylogeny analyses to provide and compare age estimates and patterns of association between these clades. Our study provides strong support for the evolution of parasitism of *Cryptosporidium* with the rise of the vertebrates about 600 million years ago (Mya). Interestingly, periods of increased diversification in *Cryptosporidium* coincides with diversification of crown mammalian and avian orders after the Cretaceous-Palaeogene (K-Pg) boundary, suggesting that adaptive radiation to new mammalian and avian hosts triggered the diversification of this parasite lineage. Despite evidence for ongoing host shifts we also found significant correlation between protozoan parasites and vertebrate hosts trees in the cophylogenetic analysis. These results help us to understand the underlying macroevolutionary mechanisms driving evolution in *Cryptosporidium* and may have important implications for the ecology, dynamics and epidemiology of cryptosporidiosis disease in humans and other animals.

Key words: Coevolution, *Cryptosporidium*, molecular clock, temporal congruence, Vertebrata.

INTRODUCTION

Coevolution occurs at many scales and is driven by interactions between species that lead to changes in the evolutionary trajectory of each interacting species. Host–parasite coevolution examples are numerous (algae and virus, Bellec *et al*. 2014; e.g. pocket gophers and chewing lice, Hafner *et al*. 1994; insects and fungi, Zhang *et al*. 2014) and shaped evolutionary theory (Anderson and May, 1982; May and Anderson, 1983). However, unresolved evolutionary histories of several parasitic groups preclude analyses of coevolutionary relationships and the timing of events of the intimate relationship with their hosts.

The evolutionary relationships and time of divergence among major Protozoa groups are contentious (Adl *et al*. 2007). Although all members of apicomplexans are parasitic and share specific features related to parasitism (e.g. an apical secretory structure mediating locomotion and cellular invasion), its extreme radiation (>6000 species known), adaptation to different niches in higher level eukaryotes (targeted hosts), lack of distinguishable morphological characters, genomic variation and complex life cycles involving multiple stages of infections make it difficult to recover deep evolutionary history and ancestry (Javaux *et al*. 2001; Templeton *et al*. 2004; Keeling *et al*. 2005; Ginger, 2006; Adl *et al*. 2007; Kuo *et al*. 2008; Wasmuth *et al*. 2009; De Baets and Littlewood, 2015). Compelling evidence, however, has progressively emerging and our knowledge of the diversity, origin and evolution of parasitic protists have benefited from molecular methods (Gilabert and Wasmuth, 2013; Sierra *et al*. 2016).

One of the most important infectious diseases in vertebrates is caused by the Apicomplexa protozoan *Cryptosporidium*. Different species of this unicellular organism have been found in all living vertebrate groups with some species shared within the same taxonomic Class (e.g. a wide range of mammals including humans, sheep, goats and cattle are the hosts of *Cryptosporidium parvum*). Species of *Cryptosporidium* are morphologically indistinguishable and their identification is mainly based on molecular characterization (Xiao *et al*. 1999; Fayer, 2010). The phylogenetic position has also been debated with the genus placed within the coccidian clade initially, whereas recent molecular studies confirmed a close affinity to the gregarines (Carreno *et al*. 1999; Barta and Thompson, 2006).

To the best of our knowledge no molecular clock analysis has been applied to establish the timeline of *Cryptosporidium* evolution and test the congruence of its time of diversification to the origin of major groups of host vertebrates. The evolutionary

* Corresponding author: Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Private Bag, 11 222, Palmerston North 4442, New Zealand. E-mail: j.c.garciaramirez@massey.ac.nz

doi:10.1017/S0031182016001323

origins and extent of host–parasite interactions can be inferred from time calibrated tree phylogenies. The symmetry in times of divergence between hosts and parasites can provide evidence of co-evolution. So, linking results that yield similar dates of divergence from dated trees of host–parasite associations at least hints that a common history of interacting lineages is shared (De Vienne *et al*. 2013). Here, we use molecular data, a number of calibration points and cophylogeny to compare temporal phylogenies and interactions between *Cryptosporidium* and their hosts in order to understand the underlying macroevolutionary mechanisms driving evolution of *Cryptosporidium* diversity. Does the origin of *Cryptosporidium* follow that of the origin of modern Vertebrata clades?

METHODS

Taxon sampling

We assembled a dataset of DNA sequences deposited in GenBank corresponding to 18S ribosomal RNA (18S), actin gene (actin) and 70 kilodalton heat shock protein (hsp70). Our sampling includes data from 27 species within the NCBI taxonomy database. Sequence data of additional Apicomplexa species were downloaded from GenBank as a close outgroup. These lineages were from groups closely related to *Cryptosporidium* (e.g. gregarines, coccidia and hematozoa) and provide appropriate context for dating analyses (Table 1). Sequences of other lineages within Alveolata (Ciliophora) were retrieved and included within the analysis. Rhizaria and Stramenopiles species were used as a known outgroup to all these taxa. We obtained two or more sequences of the same species from different sources when available to minimize systematic errors. After comparison only one sequence for each species was retained for subsequent analysis. A list of specimens and GenBank accession numbers of the sequences included in this study are presented in Table 1.

Phylogenetic analysis

Alignment of individual datasets was performed with SATé-II program v2·2·7 using MAFFT aligner and MUSCLE merger (Liu *et al*. 2012). Each gene alignment was checked by eye and further refined by hand prior to phylogenetic analysis. The substitution model was chosen in jModelTest v0·1·1 (Posada, 2008) based on the Akaike Information Criterion (Posada, 2008). Prior to concatenated analyses, single gene datasets were inspected for evidence of significant incongruence by comparing preliminary Maximum Likelihood (ML) trees using RAxML v8·2·4 (Stamatakis *et al*. 2008; Stamatakis, 2014) and a general time reversible model with gamma distribution (GTR + Γ). We observed no significant conflict among individual phylogenies and all subsequent analyses were performed with concatenated data. A 4-way partition by gene strategy was used for the concatenated analysis. The partition scheme was as follow: the fragment of the 18S rRNA and first-, second and third-codon position for the protein-coding actin and hsp70 genes. RY-coding at the third codon position was used as a partition strategy. ML analyses were implemented in RAxML using a GTR + Γ model with bootstrapping automatically stopped employing the majority rule criterion. Bayesian phylogenetic analyses (BA) were implemented in MrBayes v3·2·6 (Ronquist and Huelsenbeck, 2003; Ronquist *et al*. 2012) using 10 million generations sampled every 5000th generation, a burn in of 10%, and GTR + Γ + I model of evolution. Convergence and mixing were assessed using Tracer v1·6 (http://tree.bio.ed.ac.uk/software/tracer/) by examining log-likelihood values across generations and ensuring that post-burn-in samples yielded an effective sample size (ESS) of >200 for all parameters. RAxML and MrBayes analyses were performed via the CIPRES portal (Miller *et al*. 2010). Trees were viewed using FigTree v1·4·2 (http://tree.bio.ed.ac.uk/software/figtree/).

Molecular dating of Cryptosporidium

Divergence times were estimated in BEAST v1·8·0 (Drummond and Rambaut, 2007) using the dataset partitioned as for the phylogenetics analyses and an uncorrelated relaxed Bayesian clock with rates among branches distributed according to a log-normal distribution (Drummond *et al*. 2006). A relaxed clock model can account the variation in substitution rates among lineages (Thorne *et al*. 1998) while a lognormal distribution accommodates greater flexibility regarding a cladogenetic event (Ho and Phillips, 2009). A Birth-Death process was implemented for the speciation model (Rooney, 2004). The XML file was generated using BEAUti v1·8·0 (Drummond *et al*. 2012) with subsequent modifications by hand. The following dates and calibration priors were used according to mean date estimations in Parfrey *et al*. (2011). The root prior had a normal distribution of 1365–1577 Mya (95% range) and Rhizaria a normal distribution of 1017–1256 Mya (95% range). For comparison, we also used other calibration constraints as found in Parfrey *et al*. (2011) and Eme *et al*. (2014). First, a normal distribution of 1110–1315 Mya (95% range) for the root prior and 816–1016 (95% range) for the time of the most common ancestor (tmrca) in Rhizaria, secondly, a prior of 1371–1626 Mya (95% range) and 983–1266 (95% range) for Rhizaria, according to analysis (*b*) and (*e*) in Parfrey *et al*.

https://doi.org/10.1017/S0031182016001323 Published online by Cambridge University Press

Table 1. Taxa, major clades, GenBank accession numbers and host range of *Cryptosporidium* species included in this study

Species	Major clade	GenBank Accession No.			Host range
		18S	actin	hsp70	
Cryptosporidium andersoni	Alveolates	KF826312		AB610481	Cattle
Cryptosporidium baileyi		L19068	AF382346	KM977645	Chicken, turkey
Cryptosporidium bovis		JX515546	AY741307	AY741306	Cattle
Cryptosporidium canis		KC445656	EU754841	AY120920	Dog
Cryptosporidium cuniculus		HQ397716	GU327783	KC157562	Human, rabbit
Cryptosporidium erinacei		KF612324	KF612326	KF612325	Hedgehog
Cryptosporidium fayeri		KP730318	KP730322		Kangaroo
Cryptosporidium felis		KJ194110	AF382347	KM977646	Cat
Cryptosporidium fragile		JX948130			Toad
Cryptosporidium galli		HM116388	AY163901	AY168849	Finch, chicken
Cryptosporidium hominis		DQ286403	KP314262	KP314260	Human
Cryptosporidium macropodum		KP730303			Kangaroo
Cryptosporidium meleagridis		HQ917075	AF382351	JX024763	Turkey
Cryptosporidium molnari		HM243547	HM365220		Gilthead bream, European seabass
Cryptosporidium muris		EU553592	KJ746834	KJ746835	Mouse
Cryptosporidium parvum		AF112569	M86241	KC885895	Human, cattle, sheep, goat
Cryptosporidium ryanae		JN400880	FJ463206	EU410346	Cattle
Cryptosporidium scrofarum		KC481231	AB852580	JX424842	Pig
Cryptosporidium serpentis		EU553553	AF382353	AF221541	Snake, lizard
Cryptosporidium struthionis		AJ697751			Ostrich
Cryptosporidium suis		JQ936502	EF012372	DQ898164	Pig
Cryptosporidium tyzzeri		JX679086	JQ073414		Mouse
Cryptosporidium ubiquitum		KP730300	HM209377	HM485436	Deer
Cryptosporidium varanii		KM870593	AF382349	FJ429598	Green tree monitor, snake
Cryptosporidium viatorum		JX644908	JN846707	JX978273	Human
Cryptosporidium wrairi		AF115378	AF382348	AF221536	Guinea pig
Cryptosporidium xiaoi		KP004203	GQ337964	KF907826	Sheep
Ascogregarina taiwanensis		EF666482			
Mattesia geminata		AY334568			
Syncystis mirabilis		DQ176427			
Babesia gibsoni		KC461261		AB248730	
Theileria orientalis		HM538218			
Toxoplasma gondii		L24381		U85648	
Hammondia hammondi		AF096498			
Hammondia heydorni		GQ984224		DQ997572	
Paramecium tetraurelia		AB252008			
Chilodonella uncinata		AF300282	EU047828		
Thalassiosira pseudonana	Stramenopiles	HM991698			
Phaeodactylum tricornutum		EF140622	AY729845		
Aureococcus anophagefferens		U40257			
Heterosigma akashiwo		AB001287	AY729842	AY729866	
Apodachlya brachynema		AJ238663	AY729840		
Bodomorpha minima	Rhizaria	AF411276	FJ973394		
Heteromita globosa		U42447			
Rhizosphaera trigonacantha		JQ706069			
Collophidium ellipsoides		AB690557			
Acanthostaurus nordgaardi		HQ651787			

(2011), respectively. Divergence estimations based on the CIR clock model with soft- (900–1580 Mya) and hard-bound (1500–1850 Mya) calibration constraints in Eme *et al*. (2014) were also included. We combined the results of three independent runs of 40 million generations each to ensure ESS were above 200. TreeAnnotator v1·8·0 (Drummond and Rambaut, 2007) was used to combine and summarize trees files, obtain a maximum clade credibility consensus tree, and calculate 95% credibility intervals. Chains were sampled every 4000th generation and a burn-in of 10% (4 million generations) was used. Convergence and diagnostics of the Markov process were evaluated by the stability of parameter estimates across generations using Tracer v1·6 (http://tree.bio.ed.ac.uk/software/tracer/). The tree with the times of divergences and Highest Posterior Density (HPD) intervals was visualized using FigTree v1·4·2 (http://tree.bio.ed.ac.uk/software/figtree/).

https://doi.org/10.1017/S0031182016001323 Published online by Cambridge University Press

Dating of vertebrate evolution

The relationships and ages of major clades of vertebrates were based on those estimated by Hedges and Kumar (2009). For comparative analysis we also used molecular timescales for vertebrate evolution as found in Wiens (2015).

Global fit tests

Global fit analyses and tanglegram visualization were performed on ML tree analyses of *Cryptosporidium* and their hosts (Table 1). *Cytochrome b* sequence data were used to generate a ML tree for the most predominant hosts that have been documented for *Cryptosporidium* species (Xiao *et al*. 2002, 2004; Fayer, 2010; Šlapeta, 2013). Distance matrices were calculated using the 'cophenetic' and 'dist.node' commands within the 'ape' package in R (Paradis *et al*. 2004; R Development Core Team, 2014). A third rectangular matrix was generated for host-parasite links allowing multiple linkages between host and parasite species. We estimated the overall congruence between host and parasite topologies using the patristic distance matrices with the null hypothesis of independent evolution in ParaFit (Legendre *et al*. 2002). The fit between the *Cryptosporidium* and host topologies was assessed using the distance-based analysis and a 'cailliez' correction (Cailliez, 1983) with 999 permutations.

RESULTS

Phylogenetic analysis

The complete alignment of the three gene fragments contained 4653 bp comprising 1850 bp of 18S, 1056 bp of actin and 1747 bp of hsp70. Bayesian inference yielded a consensus tree that was topologically congruent with the ML tree, with ML bootstrap support and Bayesian posterior probabilities largely consistent among nodes (Fig. 1A and Supplementary Fig. S1). We identified three well-supported clades for internal groups within *Cryptosporidium* with similar levels of statistical support from ML and Bayesian analyses (Fig. 1A). The first well-supported split leads to a clade comprising only *Cryptosporidium 'struthionis'* (clade A), the second clade includes *Cryptosporidium galli, Cryptosporidium fragile, Cyptosporidium serpentis, Cryptosporidium andersoni* and *Cryptosporidium muris* (clade B) and a third large clade includes all other species (clade C).

Timing of diversification

Our study showed that the most recent common ancestor of the *Cryptosporidium* parasite lineage is found near to the Paleozoic/Proterozoic boundary about 590 (877–345) Mya (Fig. 1A) and represents a basal split to the clade composed by *C. 'struthionis'*. The estimated time for the split of the other two major clades within *Cryptosporidium* occurred during the middle Paleozoic ~368 (560–218) Mya, but clade B lineage formation was around the late Jurassic 162 (291–76) Mya whereas clade C originated during the late Paleozoic 265 (409–153) Mya. Among representatives of *Cryptosporidium* within clade C there was evidence of several relatively early lineage-splitting events since the Paleogene (Fig. 1A). Differences in divergence times for the crown *Cryptosporidium* clade reported from all other analyses are relatively small with estimated times after 400 Mya and before 700 Mya but the width of the 95% HPD intervals overlapping among interval age estimations (Supplementary Figs S2–S5).

The molecular clock based on an analysis by Hedges and Kumar (2009) showed that the most common ancestor of extant vertebrates is found around 600 Mya. The ages of the Vertebrata origin estimated by Hedges and Kumar (2009) are older than those estimated by Wiens (2015). These time trees differ in the crown age of Vertebrata by about 100 My. The phylogeny and time of divergences of the major vertebrate clades is also shown in Fig. 1B.

Global fit tests

The cophylogenetic analysis also revealed statistically significant patterns of association between hosts and parasites (Global test: ParaFitGlobal = 1·02, *P*-value = 0·01). Comparisons of host and parasite phylogenies based on distance and topology-based analyses provided support for a common macroevolutionary scenario between *Cryptosporidium* and their vertebrate hosts (Fig. 2).

DISCUSSION

Our comparison of the divergence times provides evidence for the origin of *Cryptosporidium* parasites close to the time of the most common ancestor for all vertebrates about 600 Mya. Different calibration points used in this study yield no significant differences for the root of extant *Cryptosporidium* clade. However, estimated ages for the crown group of *Cryptosporidium* are older [679 (1012–393) Mya] and younger [408 (703–180) Mya] when a CIR model and hard- and soft-bound is respectively implemented. These times of the origin of *Cryptosporidium* nevertheless overlap with interval age estimations reported for the origin of Vertebrata (Kumar and Hedges, 1998; Blair and Hedges, 2005; Erwin *et al*. 2011; Hedges *et al*. 2015). The basal split between clades B and C about 400 Mya is also congruent with the age of the Actinopterygii clade where fish species that are

https://doi.org/10.1017/S0031182016001323 Published online by Cambridge University Press

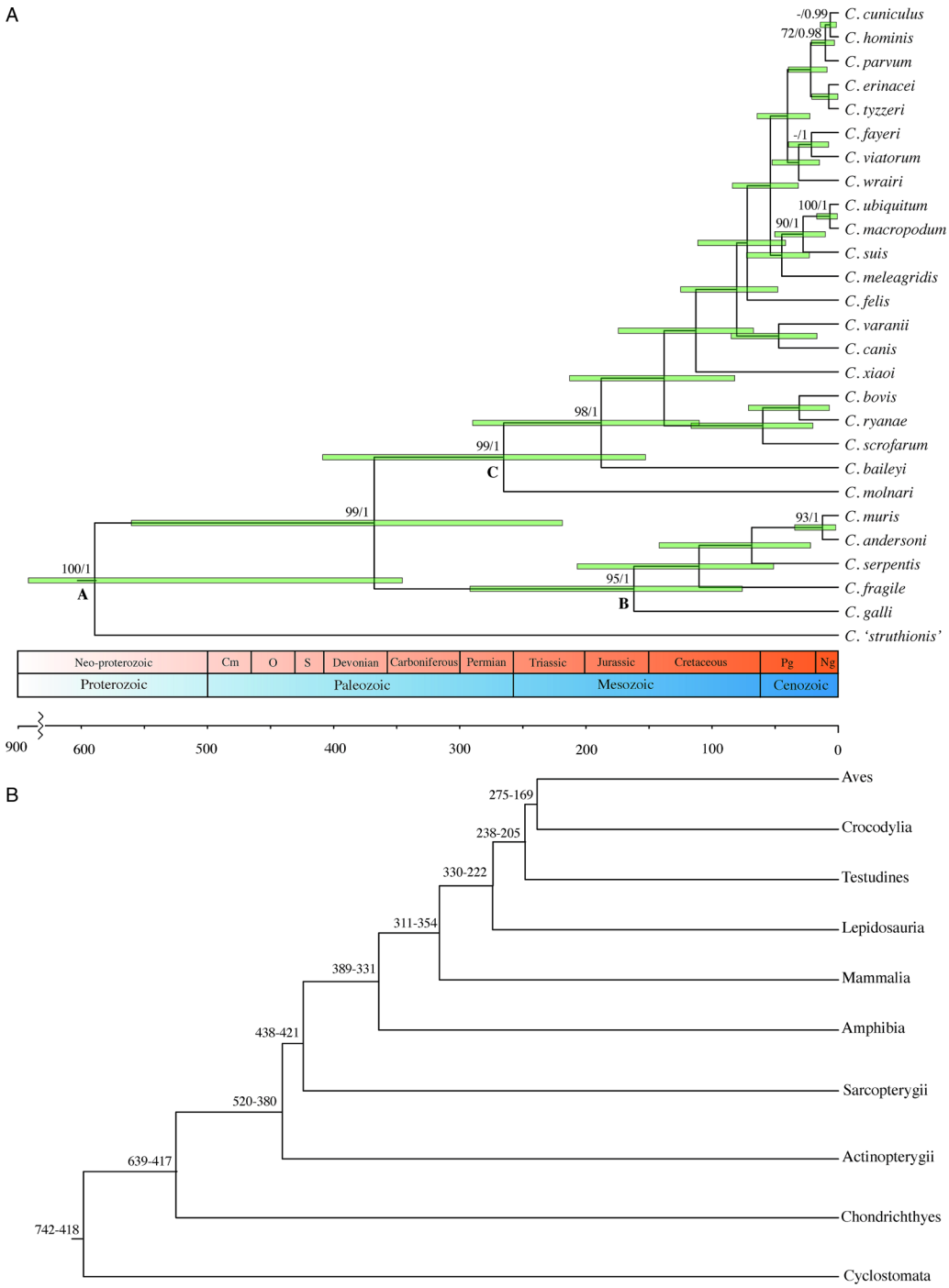

Fig. 1. (A) Chronogram of *Cryptosporidium* based on concatenated genes (18S, actin and hsp70) with a Lognormal relaxed-clock Bayesian analysis using BEAST. Age constraints were established by a root prior with a normal distribution of 1365–1577 Mya (95% range) and Rhizaria a normal distribution of 1017–1256 Mya (95% range). For each node the estimate time of divergence and 95% Highest Posterior Density (HPD) intervals are shown. The timescale is in millions of years ago (Mya) and geological eras and periods are indicated where Ng (Neogene), Pg (Paleogene), S (Silurian), O (Ordovician) and Cm (Cambrian). Bootstrap support over 70% and Bayesian posterior probabilities over 0.9 are found above each branch. Letters below the nodes refer to clades discussed in the text. A complete figure including all species analysed in this study is found in Supplementary Figure S1. (B) A timetree representing temporal patterns of diversification in major lineages of vertebrates. Topology and divergence dates are consensus estimates derived from Hedges and Kumar (2009) and Wiens (2015). Confidence intervals among vertebrate clades are found in each branch following estimates from Blair and Hedges (2005) and Kumar and Hedges (1998). Confidence interval for the origin of Vertebrata includes minimum and maximum age estimations from both studies.

hosts to *Cryptosporidium molnari* belong to. Analysis of the dated molecular phylogenies suggests that the origin of the clade C, which infects mainly mammalian hosts, is concordant with the age of the stem group of mammals during the Triassic (Close *et al*. 2015). Yet much of the taxonomic diversity

https://doi.org/10.1017/S0031182016001323 Published online by Cambridge University Press

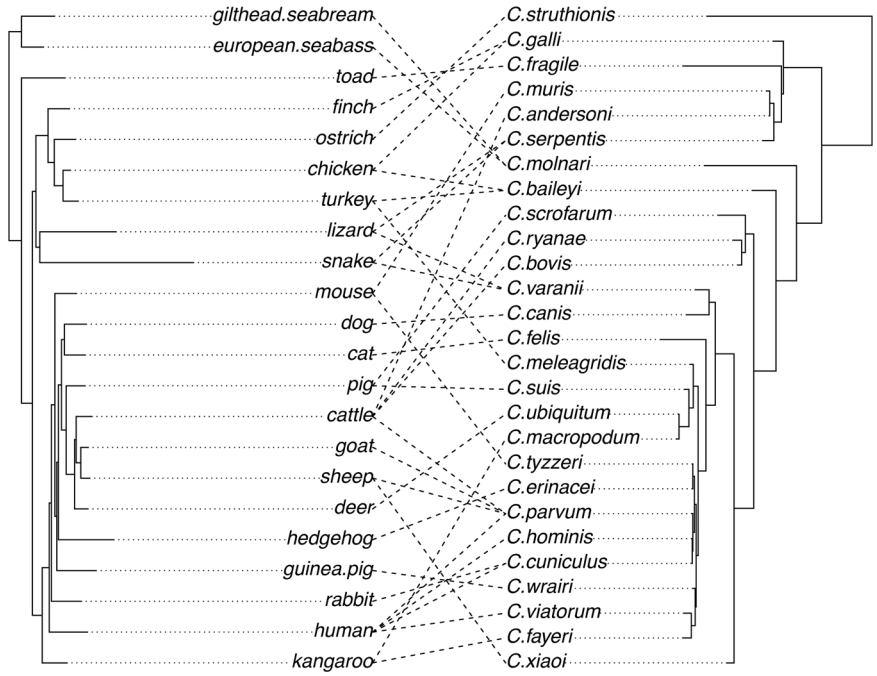

Fig. 2. Tanglegram depicting the host–parasite relationships between *Cryptosporidium* species (right) and their most dominant vertebrate hosts (left). Phylogenies were reconstructed with Maximum Likelihood (ML) analysis using concatenated data for parasites (18S, actin and hsp70) and a single mtDNA gene (cyt*b*) for hosts.

of *Cryptosporidium* originated in the Cretaceous, as did most of the terrestrial vertebrates groups (Cooper and Penny, 1997; Kumar and Hedges, 1998). Taxonomic and ecological diversity in *Cryptosporidium* appears to have evolved during the Cretaceous and provided a launching pad for later diversification during the Tertiary period when mammalian and avian orders diversified after the K–Pg event (Dos Reis *et al*. 2012; O'Leary *et al*. 2013; Jarvis *et al*. 2014; Claramunt and Cracraft, 2015; Prum *et al*. 2015). In this respect the evolution of these parasites mirrors the evolution of vertebrates, primarily in terms of the diversification of terrestrial Eutheria and Metatheria mammals and Palaeognathae and Neognathae birds (e.g. Jetz *et al*. 2012; Jarvis *et al*. 2014; Close *et al*. 2015). Our analyses also find support for the evolution of *Cryptosporidium hominis* with our human ancestors. The split between *C. hominis* and *C. cuniculus* around 6 (1·4–14) Mya suggests an approximate date concordant with our hominini ancestor likely tracing the evolution of *C. hominis* parasite back to that speciation event (Langergraber *et al*. 2012).

The age congruencies regarding the coevolution of *Cryptosporidium* and vertebrates from our estimation of divergence times are supported by the global fit test of host-parasite cophylogenetic pattern. The cophylogenetic statistical analysis indicates a predominance of coevolution compared with host shifting despite some parasites infecting multiple hosts. Some *Cryptosporidium* species seem to be host-restricted to a single host (e.g. *C. viatorum* has been only found in humans) but others are distributed across different hosts (e.g. *C. parvum* is found in humans, cattle, sheep, goats) sometimes achieving high prevalence in one or more hosts (Xiao *et al*. 2002, 2004; Fayer, 2010; Cacciò and Widmer, 2013; Šlapeta, 2013). *Cryptosporidium* species infecting closely related hosts within some subgroups is especially common within clade C. For instance, *C. parvum, C. hominis* and *C. cuniculus* seem to arise owing to movement and specialization to new mammal hosts (e.g. Koehler *et al*. 2014). These species are not sufficiently specialized to individual hosts to prevent gene flow; therefore it is likely that shifting occurs because there are not ecological barriers for their populations to disperse among different closely related hosts. Such host shifting could be involved in coevolution of resistance factors by the host populations (Ricklefs *et al*. 2014) but finer resolution analysis, preferably using whole-genome sequences over shorter timescales, are likely required to resolve these parasite-host population level questions.

Host shifting through different host-vertebrate combinations might indicate that the diversity of *Cryptosporidium* parasites has not been determined yet. Numerous diverse isolates have been characterized probably encompassing more species than those formally described so far (e.g. Alvarez-Pellitero *et al*. 2004; Li *et al*. 2015; Ryan *et al*. 2015). For example, the still undescribed strain *Cryptosporidium 'struthionis'* has been isolated from ostrich, yet close relatives strains have been found in coprolites of moa (Wood *et al*. 2013) and free-living in

https://doi.org/10.1017/S0031182016001323 Published online by Cambridge University Press

tidal-flat (Wilms *et al*. 2006) and ballast water (Pagenkopp *et al*. 2016). *Cryptosporidium 'struthionis'* is on a relatively long branch with seemingly phylogenetically deep origins. This long-branch would probably be broken with additional taxon sampling and sequence data (Bergsten, 2005; Slack *et al*. 2007). Future taxonomic work will impact our understanding of *Cryptosporidium* evolution dramatically and will stimulate comparative studies to address the growing number of questions regarding the evolution of protozoan parasites.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at http://dx.doi.org/10.1017/S0031182016001323.

ACKNOWLEDGEMENTS

The first author (JCGR) would like to thank to the New Zealand Ministry of Health for support. ^{m}EpiLab members provided useful discussions on different stages of the study. We are grateful with two anonymous reviewers for provided helpful comments that improved this manuscript.

FINANCIAL SUPPORT

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

REFERENCES

Adl, S. M., Leander, B. S., Simpson, A. G. B., Archibald, J. M., Anderson, O. R., Bass, D., Bowser, S. S., Brugerolle, G., Farmer, M. A., Karpov, S., Kolisko, M., Lane, C. E., Lodge, D. J., Mann, D. G., Meisterfeld, R., Mendoza, L., Moestrup, Ø., Mozley-Standridge, S. E., Smirnov, A. V. and Spiegel, F. (2007). Diversity, nomenclature, and taxonomy of Protists. *Systematic Biology* **56**, 684–689.

Alvarez-Pellitero, P., Quiroga, M. I., Sitja-Bobadilla, A., Redondo, M. J., Palenzuela, O., Vazquez, P. and Nieto, J. M. (2004). *Cryptosporidium scophthalmi* n. sp. (Apicomplexa: Cryptosporidiidae) from cultured turbot *Scophthalmus maximus*. Light and electron microscope description and histopathological study. *Diseases of Aquatic Organisms* **62**, 133–145.

Anderson, R. M. and May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology* **85**, 411–426.

Barta, J. R. and Thompson, R. C. A. (2006). What is *Cryptosporidium*? Reappraising its biology and phylogenetic affinities. *Trends in Parasitology* **22**, 463–468.

Bellec, L., Clerissi, C., Edern, R., Foulon, E., Simon, N., Grimsley, N. and Desdevises, Y. (2014). Cophylogenetic interactions between marine viruses and eukaryotic picophytoplankton. *BMC Evolutionary Biology* **14**, 1–13.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* **21**, 163–193.

Blair, J. E. and Hedges, S. B. (2005). Molecular phylogeny and divergence times of Deuterostome animals. *Molecular Biology and Evolution* **22**, 2275–2284.

Cacciò, S. M. and Widmer, G. (2013). *Cryptosporidium: Parasite and Disease*. Springer Science & Business Media, Vienna.

Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika* **48**, 305–308.

Carreno, R. A., Matrin, D. S. and Barta, J. R. (1999). *Cryptosporidium* is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences. *Parasitology Research* **85**, 899–904.

Claramunt, S. and Cracraft, J. (2015). A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances* **1**, e1501005.

Close, R. A., Friedman, M., Lloyd, G. T. and Benson, R. B. J. (2015). Evidence for a Mid-Jurassic adaptive radiation in mammals. *Current Biology* **25**, 2137–2142.

Cooper, A. and Penny, D. (1997). Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science* **275**, 1109–1113.

De Baets, K. and Littlewood, D. T. J. (2015). The importance of fossils in understanding the evolution of parasites and their vectors. In *Advances in Parasitology* (ed. Kenneth De, B. & Littlewood, D. T. J.), pp. 1–51. Academic Press, London, UK.

De Vienne, D. M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M. E. and Giraud, T. (2013). Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist* **198**, 347–385.

Dos Reis, M., Inoue, J., Hasegawa, M., Asher, R. J., Donoghue, P. C. J. and Yang, Z. (2012). Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society of London B: Biological Sciences* **279**, 3491–3500.

Drummond, A. J. and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 1–8.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *Plos Biology* **4**, e88.

Drummond, A. J., Suchard, M. A., Xie, D. and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1·7. *Molecular Biology and Evolution* **29**, 1969–1973.

Eme, L., Sharpe, S. C., Brown, M. W. and Roger, A. J. (2014). On the age of Eukaryotes: evaluating evidence from fossils and molecular clocks. In *The Origin and Evolution of Eukaryotes* (ed. Keeling, P. J. & Koonin, E. V.), pp. 165–180. Cold Spring Harbor Laboratory Press, New York.

Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D. and Peterson, K. J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**, 1091–1097.

Fayer, R. (2010). Taxonomy and species delimitation in *Cryptosporidium*. *Experimental Parasitology* **124**, 90–97.

Gilabert, A. and Wasmuth, J. D. (2013). Unravelling parasitic nematode natural history using population genetics. *Trends in Parasitology* **29**, 438–448.

Ginger, M. L. (2006). Niche metabolism in parasitic protozoa. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 101–118.

Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, J. W. and Nadler, S. A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* **265**, 1087–1090.

Hedges, S. B. and Kumar, S. (2009). *The Timetree of Life*. Oxford University Press, New York.

Hedges, S. B., Marin, J., Suleski, M., Paymer, M. and Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution* **32**, 835–845.

Ho, S. Y. W. and Phillips, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58**, 367–380.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., Da Fonseca, R. R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M. S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M. *et al*. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331.

Javaux, E. J., Knoll, A. H. and Walter, M. R. (2001). Morphological and ecological complexity in early eukaryotic ecosystems. *Nature* **412**, 66–69.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. and Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature* **491**, 444–448.

Keeling, P. J., Burger, G., Durnford, D. G., Lang, B. F., Lee, R. W., Pearlman, R. E., Roger, A. J. and Gray, M. W. (2005). The tree of eukaryotes. *Trends in Ecology & Evolution* **20**, 670–676.

Koehler, A. V., Whipp, M. J., Haydon, S. R. and Gasser, R. B. (2014). *Cryptosporidium cuniculus* - new records in human and kangaroo in Australia. *Parasites & Vectors* **7**, 492.

Kumar, S. and Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920.

Kuo, C.-H., Wares, J. P. and Kissinger, J. C. (2008). The apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. *Molecular Biology and Evolution* **25**, 2689–2698.

Langergraber, K. E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J. C., Muller, M. N., Robbins, M. M., Schubert, G., Stoinski, T. S., Viola, B., Watts, D., Wittig, R. M., Wrangham, R. W., Zuberbühler, K., Pääbo, S. and Vigilant, L. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences* **109**, 15716–15721.

Legendre, P., Desdevises, Y. and Bazin, E. (2002). A statistical test for host–parasite coevolution. *Systematic Biology* **51**, 217–234.

Li, X., Pereira, M. D. G. C., Larsen, R., Xiao, C., Phillips, R., Striby, K., Mccowan, B. and Atwill, E. R. (2015). *Cryptosporidium rubeyi* n. sp. (Apicomplexa: Cryptosporidiidae) in multiple *Spermophilus* ground squirrel species. *International Journal for Parasitology: Parasites and Wildlife* **4**, 343–350.

Liu, K., Warnow, T. J., Holder, M. T., Nelesen, S. M., Yu, J., Stamatakis, A. P. and Linder, C. R. (2012). SATé-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology* **61**, 90–106.

May, R. M. and Anderson, R. M. (1983). Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **219**, 281–313.

Miller, M. A., Pfeiffer, W. and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. pp. 1–8. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA.

O'leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., Goldberg, S. L., Kraatz, B. P., Luo, Z.-X., Meng, J., Ni, X., Novacek, M. J., Perini, F. A., Randall, Z. S., Rougier, G. W., Sargis, E. J., Silcox, M. T., Simmons, N. B., Spaulding, M., Velazco, P. M., Weksler, M., Wible, J. R. and Cirranello, A. L. (2013). The placental mammal ancestor and the post–K-Pg radiation of placentals. *Science* **339**, 662–667.

Pagenkopp, K. M., Fleischer, R. C., Carney, K. J., Holzer, K. K. and Ruiz, G. M. (2016). Amplicon-based pyrosequencing reveals high diversity of protistan parasites in ships' ballast water: implications for biogeography and infectious diseases. *Microbial Ecology* **71**, 530–442.

Paradis, E., Claude, J. and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290.

Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. and Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences* **108**, 13624–13629.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253–1256.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M. and Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ricklefs, R. E., Outlaw, D. C., Svensson-Coelho, M., Medeiros, M. C. I., Ellis, V. A. and Latta, S. (2014). Species formation by host shifting in avian malaria parasites. *Proceedings of the National Academy of Sciences* **111**, 14816–14821.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. and Huelsenbeck, J. P. (2012). MrBayes 3·2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539–542.

Rooney, A. P. (2004). Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in Apicomplexans. *Molecular Biology and Evolution* **21**, 1704–1711.

Ryan, U., Paparini, A., Tong, K., Yang, R., Gibson-Kueh, S., O'hara, A., Lymbery, A. and Xiao, L. (2015). *Cryptosporidium huwi* n. sp. (Apicomplexa: Eimeriidae) from the guppy (*Poecilia reticulata*). *Experimental Parasitology* **150**, 31–35.

Sierra, R., Cañas-Duarte, S. J., Burki, F., Schwelm, A., Fogelqvist, J., Dixelius, C., González-García, L. N., Gile, G. H., Slamovits, C. H., Klopp, C., Restrepo, S., Arzul, I. and Pawlowski, J. (2016). Evolutionary origins of Rhizarian parasites. *Molecular Biology and Evolution* **33**, 980–983.

Slack, K. E., Delsuc, F., Mclenachan, P. A., Arnason, U. and Penny, D. (2007). Resolving the root of the avian mitogenomic tree by breaking up long branches. *Molecular Phylogenetics and Evolution* **42**, 1–13.

Šlapeta, J. (2013). Cryptosporidiosis and *Cryptosporidium* species in animals and humans: a thirty colour rainbow? *International Journal for Parasitology* **43**, 957–970.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.

Stamatakis, A., Hoover, P. and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**, 758–771.

Templeton, T. J., Iyer, L. M., Anantharaman, V., Enomoto, S., Abrahante, J. E., Subramanian, G. M., Hoffman, S. L., Abrahamsen, M. S. and Aravind, L. (2004). Comparative analysis of Apicomplexa and genomic diversity in Eukaryotes. *Genome Research* **14**, 1686–1695.

Thorne, J. L., Kishino, H. and Painter, I. S. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15**, 1647–1657.

Wasmuth, J., Daub, J., Peregrín-Alvarez, J. M., Finney, C. a. M. and Parkinson, J. (2009). The origins of apicomplexan sequence innovation. *Genome Research* **19**, 1202–1213.

Wiens, J. J. (2015). Explaining large-scale patterns of vertebrate diversity. *Biology Letters* **11**, 1–4.

Wilms, R., Sass, H., Köpke, B., Köster, J., Cypionka, H. and Engelen, B. (2006). Specific bacterial, archaeal, and Eukaryotic communities in tidal-flat sediments along a vertical profile of several meters. *Applied and Environmental Microbiology* **72**, 2756–2764.

Wood, J. R., Wilmshurst, J. M., Rawlence, N. J., Bonner, K. I., Worthy, T. H., Kinsella, J. M. and Cooper, A. (2013). A megafauna's microfauna: Gastrointestinal parasites of New Zealand's extinct moa (Aves: Dinornithiformes). *PLoS ONE* **8**, e57315.

Xiao, L., Escalante, L., Yang, C., Sulaiman, I., Escalante, A. A., Montali, R. J., Fayer, R. and Lal, A. A. (1999). Phylogenetic analysis of *Cryptosporidium* parasites based on the small-subunit rRNA gene locus. *Applied and Environmental Microbiology* **65**, 1578–1583.

Xiao, L., Sulaiman, I. M., Ryan, U. M., Zhou, L., Atwill, E. R., Tischler, M. L., Zhang, X., Fayer, R. and Lal, A. A. (2002). Host adaptation and host–parasite co-evolution in *Cryptosporidium*: implications for taxonomy and public health. *International Journal for Parasitology* **32**, 1773–1785.

Xiao, L., Fayer, R., Ryan, U. and Upton, S. J. (2004). *Cryptosporidium* taxonomy: recent advances and implications for public health. *Clinical Microbiology Reviews* **17**, 72–97.

Zhang, Y., Zhang, S., Li, Y., Ma, S., Wang, C., Xiang, M., Liu, X., An, Z., Xu, J. and Liu, X. (2014). Phylogeography and evolution of a fungal–insect association on the Tibetan Plateau. *Molecular Ecology* **23**, 5337–5355.