TITLE: A novel metabarcoding diagnostic tool to explore protozoan haemoparasite diversity in mammals: a proof-of-concept study using canines from the tropics

AUTHORS: Lucas G Huggins*[1], Anson V Koehler[1], Dinh Ng-Nguyen[2], Stephen Wilcox[3], Bettina Schunack[4], Tawin Inpankaew[5], Rebecca J Traub[1]

AFFILIATION:

[1]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, VIC 3052, Australia

[2]Faculty of Animal Sciences and Veterinary Medicine, Tay Nguyen University, Dak Lak 630000, Vietnam
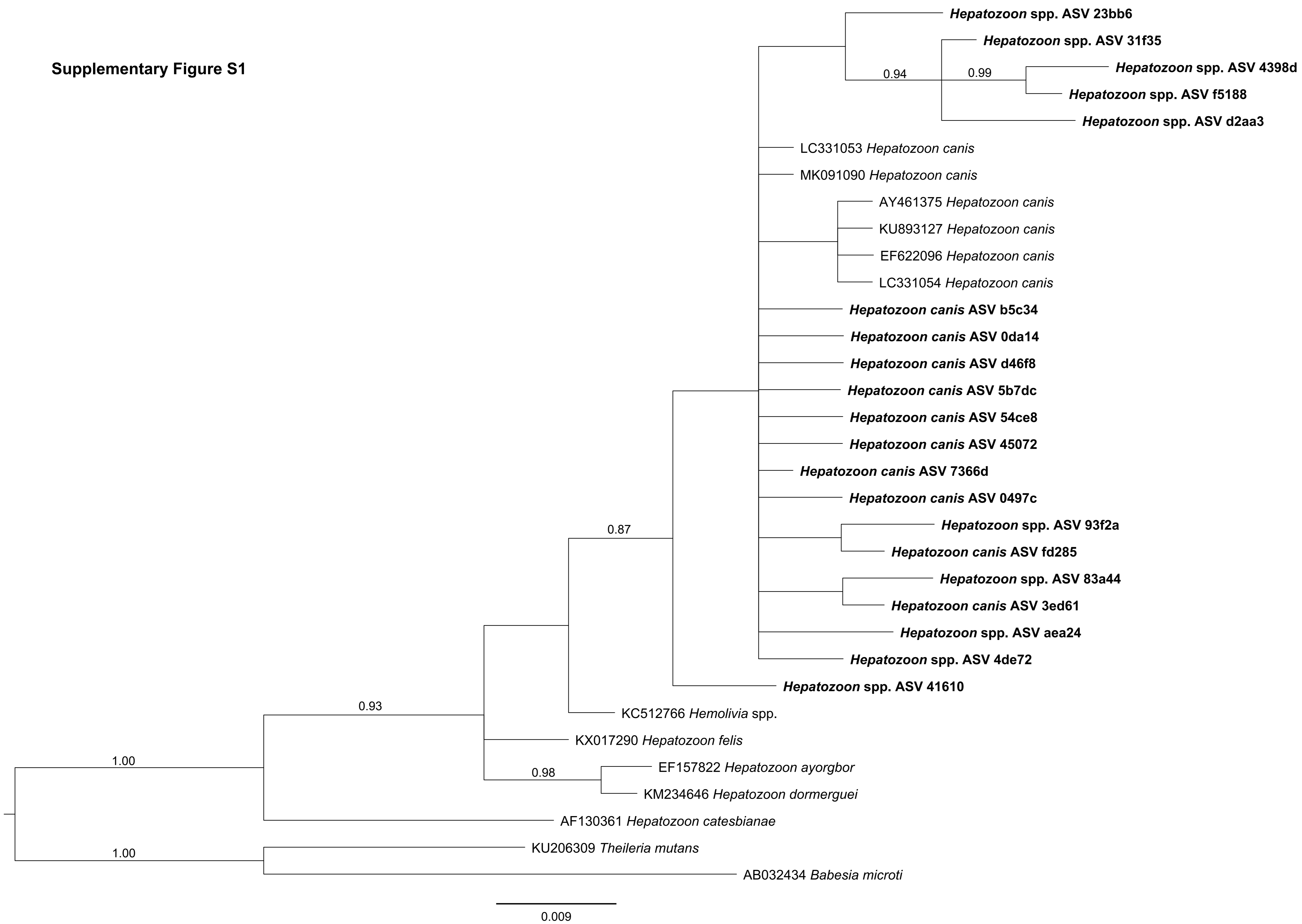
[3]Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia

[4]Bayer Animal Health GmbH, Leverkusen, Germany
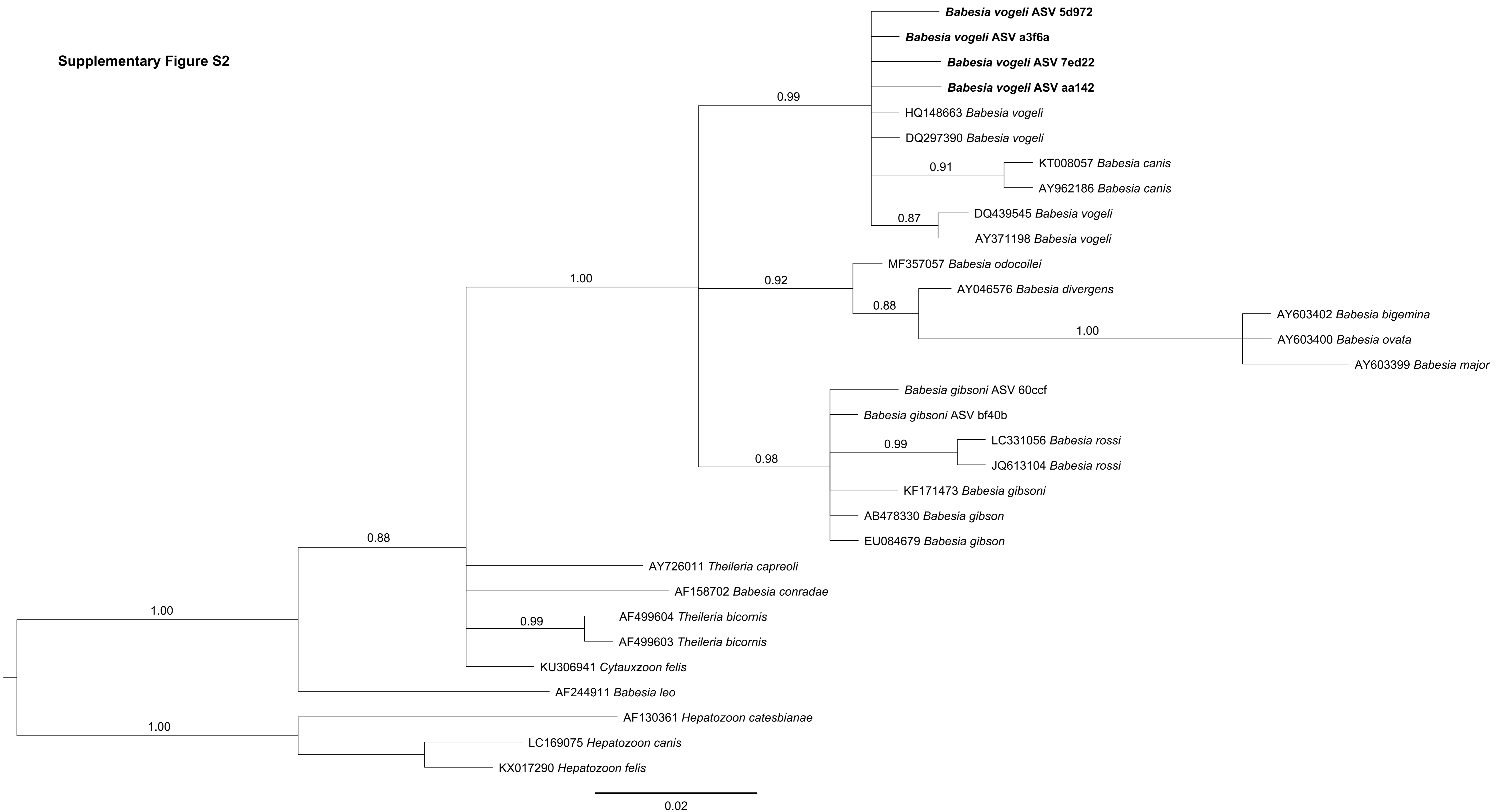
[5]Faculty of Veterinary Medicine, Kasetsart University, Bangkok 10900, Thailand

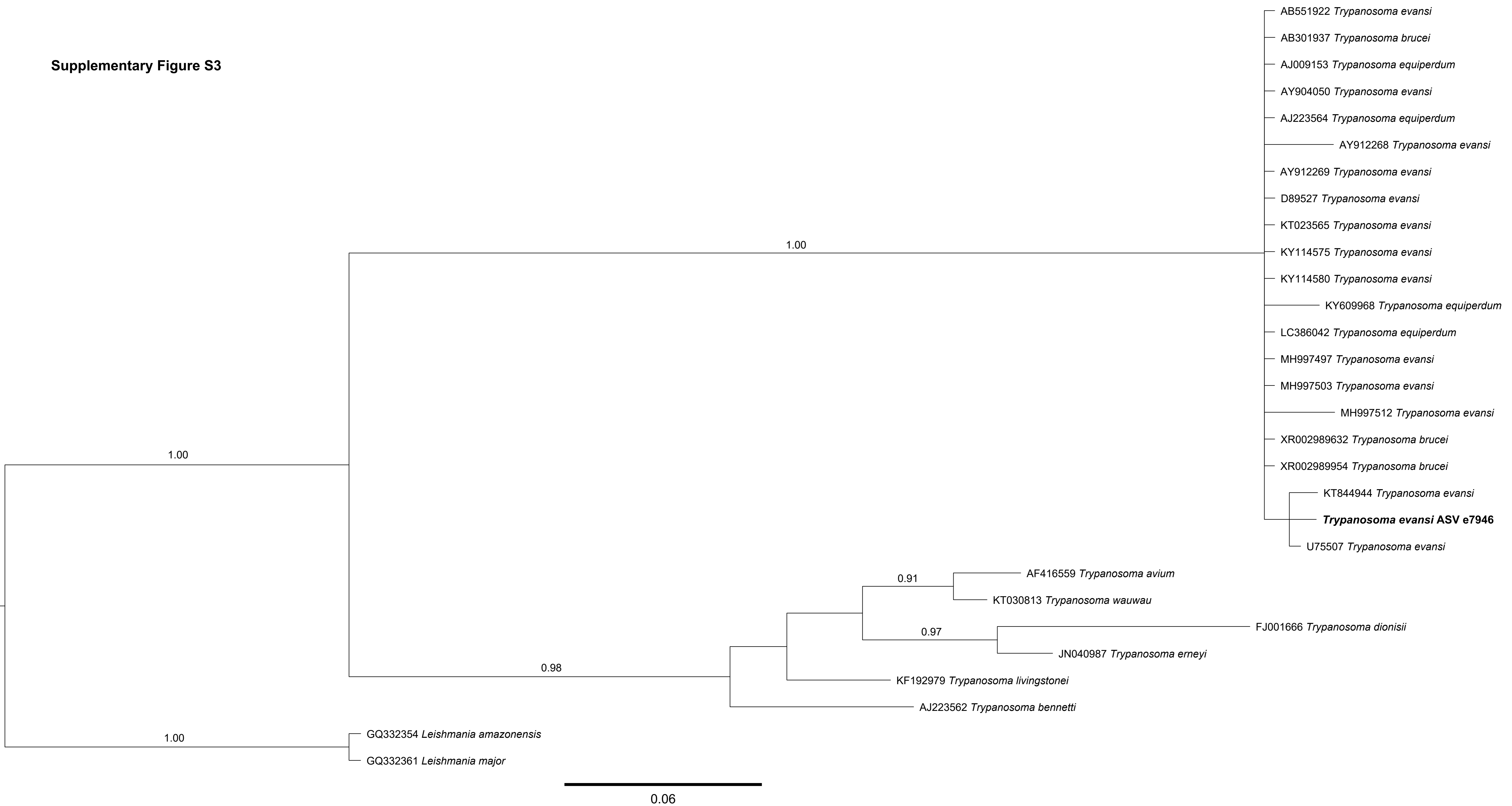*Correspondence: lghuggins@student.unimelb.edu.au

Supplementary Figure S1

Hepatozoon spp. ASV 23bb6

Hepatozoon spp. ASV 31f35

Hepatozoon spp. ASV 4398d

Hepatozoon spp. ASV f5188

Hepatozoon spp. ASV d2aa3

0.94

0.99

LC331053 *Hepatozoon canis*

MK091090 *Hepatozoon canis*

AY461375 *Hepatozoon canis*

KU893127 *Hepatozoon canis*

EF622096 *Hepatozoon canis*

LC331054 *Hepatozoon canis*

Hepatozoon canis ASV b5c34

Hepatozoon canis ASV 0da14

Hepatozoon canis ASV d46f8

Hepatozoon canis ASV 5b7dc

Hepatozoon canis ASV 54ce8

Hepatozoon canis ASV 45072

Hepatozoon canis ASV 7366d

Hepatozoon canis ASV 0497c

Hepatozoon spp. ASV 93f2a

Hepatozoon canis ASV fd285

Hepatozoon spp. ASV 83a44

Hepatozoon canis ASV 3ed61

Hepatozoon spp. ASV aea24

Hepatozoon spp. ASV 4de72

Hepatozoon spp. ASV 41610

0.87

KC512766 *Hemolivia* spp.

KX017290 *Hepatozoon felis*

EF157822 *Hepatozoon ayorgbor*

KM234646 *Hepatozoon dormerguei*

0.98

0.93

AF130361 *Hepatozoon catesbianae*

1.00

KU206309 *Theileria mutans*

AB032434 *Babesia microti*

1.00

0.009

Supplementary Figure S2

Babesia vogeli ASV 5d972
Babesia vogeli ASV a3f6a
Babesia vogeli ASV 7ed22
Babesia vogeli ASV aa142
HQ148663 *Babesia vogeli*
DQ297390 *Babesia vogeli*
KT008057 *Babesia canis*
AY962186 *Babesia canis*
DQ439545 *Babesia vogeli*
AY371198 *Babesia vogeli*
MF357057 *Babesia odocoilei*
AY046576 *Babesia divergens*
AY603402 *Babesia bigemina*
AY603400 *Babesia ovata*
AY603399 *Babesia major*
Babesia gibsoni ASV 60ccf
Babesia gibsoni ASV bf40b
LC331056 *Babesia rossi*
JQ613104 *Babesia rossi*
KF171473 *Babesia gibsoni*
AB478330 *Babesia gibson*
EU084679 *Babesia gibson*
AY726011 *Theileria capreoli*
AF158702 *Babesia conradae*
AF499604 *Theileria bicornis*
AF499603 *Theileria bicornis*
KU306941 *Cytauxzoon felis*
AF244911 *Babesia leo*
AF130361 *Hepatozoon catesbianae*
LC169075 *Hepatozoon canis*
KX017290 *Hepatozoon felis*

0.99
0.91
0.87
1.00
0.92
0.88
1.00
0.98
0.99
0.88
1.00
0.99
1.00

0.02

Supplementary Figure S3

AB551922 *Trypanosoma evansi*
AB301937 *Trypanosoma brucei*
AJ009153 *Trypanosoma equiperdum*
AY904050 *Trypanosoma evansi*
AJ223564 *Trypanosoma equiperdum*
AY912268 *Trypanosoma evansi*
AY912269 *Trypanosoma evansi*
D89527 *Trypanosoma evansi*
KT023565 *Trypanosoma evansi*
KY114575 *Trypanosoma evansi*
KY114580 *Trypanosoma evansi*
KY609968 *Trypanosoma equiperdum*
LC386042 *Trypanosoma equiperdum*
MH997497 *Trypanosoma evansi*
MH997503 *Trypanosoma evansi*
MH997512 *Trypanosoma evansi*
XR002989632 *Trypanosoma brucei*
XR002989954 *Trypanosoma brucei*
KT844944 *Trypanosoma evansi*
Trypanosoma evansi ASV e7946
U75507 *Trypanosoma evansi*
AF416559 *Trypanosoma avium*
KT030813 *Trypanosoma wauwau*
FJ001666 *Trypanosoma dionisii*
JN040987 *Trypanosoma erneyi*
KF192979 *Trypanosoma livingstonei*
AJ223562 *Trypanosoma bennetti*
GQ332354 *Leishmania amazonensis*
GQ332361 *Leishmania major*

1.00
0.91
0.97
0.98
1.00
1.00
1.00

0.06

Supplementary Figure S4

Supplementary file 1: Figure S1. *Hepatozoon* **spp. phylogenetic analysis by Bayesian inference (BI) of a 130 bp region of the 18S ribosomal RNA (rRNA) gene targeted by our apicomplexan primers.** ASVs assigned as *Hepatozoon* spp. or *Hepatozoon canis* generated from our NGS data are in bold and identified by a 5-character string code. Other sequences are references taken from GenBank, including a variety of *H. canis* sequences from different geographical regions, four closely related *Hepatozoon* species and two piroplasmid species as an outgroup. Posterior probabilities (> 0.85) are indicated on branches. The scale-bar represents the number of substitutions per site. Our *H. canis* classified ASVs cluster within *H. canis* reference sequences. **Figure S2.** *Babesia* **spp. phylogenetic analysis by Bayesian inference (BI) of a 130 bp region of the 18S ribosomal RNA (rRNA) gene targeted by our apicomplexan primers.** ASVs assigned as *Babesia vogeli* or *Babesia gibsoni* generated from our NGS data are in bold and identified by a 5-character string code. Other sequences are references taken from GenBank, including *B. vogeli* and *B. gibsoni* sequences from different geographical regions. Three *Hepatozoon* species are used as an outgroup, whilst *Theileria* spp. and *Cytauxzoon* spp. are used for comparison. Posterior probabilities (> 0.85) are indicated on branches. The scale-bar represents the number of substitutions per site. Our *B. vogeli* and *B. gibsoni* classified ASVs both cluster next to *B. vogeli* and *B. gibsoni* reference sequences, whilst closely related *Babesia canis* and *Babesia rossi* show a significant number of nucleotide substitutions from their relevant clusters. **Figure S3.** *Trypanosoma evansi* **phylogenetic analysis by Bayesian inference (BI) of a 170 bp region of the 18S ribosomal RNA (rRNA) gene targeted by our Kinetoplastidae primers.** The ASV assigned as *T. evansi* from our NGS data is in bold and identified by a 5-character string code, other sequences are references taken from GenBank, including *T. evansi* and the closely related *Trypanosoma equiperdum* and *Trypanosoma brucei* from different geographical regions. Two *Leishmania* species are used as an outgroup, whilst a range of more distantly related *Trypanosoma* are used for comparison. Posterior probabilities (> 0.85) are indicated on branches. The scale-bar represents the number of substitutions per site. Our *T. evansi* classified ASV, clusters next to *T. evansi*, *T. equiperdum* and *T. brucei* reference sequences, the latter two are known to be indistinguishable at the 18S rRNA gene, however *T. equiperdum* solely infects horses, whilst *T. brucei* is restricted to Africa, meaning only *T. evansi* can be the responsible pathogen [33,34]. **Figure S4.** *Parabodo caudatus* **phylogenetic analysis by Bayesian inference (BI) of a 170 bp region of the 18S ribosomal RNA (rRNA) gene targeted by our Kinetoplastidae primers.** The ASV assigned as *P. caudatus* from our NGS data is in bold and identified by a 5-character string code, other sequences are references taken from GenBank, including *P. caudatus* (formerly classified as *Bodo caudatus*) and the closely related *Parabodo nitrophilus* from different geographical regions [36,48]. Three *Leishmania* species are used as an outgroup. Posterior probabilities (> 0.85) are indicated on branches. The scale-bar represents the number of substitutions per site. Our *P. caudatus* classified ASV, clusters next to *P. caudatus* reference sequences with strong posterior probability support.