bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # A unique *Toxoplasma gondii* haplotype under strong selection has
2 # accompanied domestic cats in their global expansion.

3

4 Lokman Galal[a,*], Frédéric Ariey[b], Meriadeg Ar Gouilh[c,d,e], Marie-Laure Dardé[a,f], Azra
5 Hamidović[a], Franck Letourneur[g], Franck Prugnolle[h,i] and Aurélien Mercier[a,f]

6 [a] INSERM, Univ. Limoges, CHU Limoges, IRD, U1094 Tropical Neuroepidemiology, Institute of
7 Epidemiology and Tropical Neurology, GEIST, Limoges, France.

8 [b] Université de Paris, Institut Cochin, Inserm U1016, Service de Parasitologie Hôpital Cochin,
9 75014 Paris, France.

10 [c] Groupe de Recherche sur l'Adaptation Microbienne (GRAM 2.0), Normandie Université,
11 UNICAEN, UNIROUEN, EA2656, Caen, France.

12 [d] Laboratoire de Virologie, Centre Hospitalo-Universitaire, Caen, France.

13 [e] Institut de Recherche pour le Développement (IRD), Maladies Infectieuses et vecteurs:
14 Ecologie, Génétique, Evolution et Contrôle (MIVEGEC) (Université de Montpellier - IRD 224 -
15 CNRS 5290), Montpellier, France

16 [f] Centre National de Référence (CNR) Toxoplasmose/Toxoplasma Biological Center (BRC),
17 Centre Hospitalier-Universitaire Dupuytren, Limoges, France.

18 [g] Plate-Forme Séquençage et Génomique, Institut Cochin, Inserm U1016, Université de Paris,
19 22 rue Méchain, 75014 Paris, France.

20 [h] Laboratory MIVEGEC, University of Montpellier, CNRS, IRD, 900 rue Jean François Breton,
21 34090 Montpellier, France.

22 [i] CREES, Centre de Recherches en Écologie et Évolution de la Santé, Montpellier, France.

23

24 *Corresponding author
25 Lokman Galal
26 INSERM, Univ. Limoges, CHU Limoges, IRD, U1094 Tropical Neuroepidemiology, Institute of
27 Epidemiology and Tropical Neurology, GEIST, Limoges, France
28 Faculté de Médecine, Université de Limoges, Limoges 87025, France
29 Tel: +33 (0) 5 19 56 42 64
30 Fax: +33 (0) 5 55 43 58 21
31 E-mail: lokmanmagalal@gmail.com
32

33

34

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

35 ## ABSTRACT

36 *Toxoplasma gondii* is a cyst-forming apicomplexan parasite of virtually all warm-
37 blooded species, with all true cats (Felidae) as definitive hosts. It is the etiologic agent of
38 toxoplasmosis, a disease causing substantial public health burden worldwide. Its wide range
39 of host species and its global occurrence probably complicate the study of its evolutionary
40 history, and conflicting scenarios have been proposed to explain its current global
41 distribution. In this study, we analyse a global set of 156 genomes (including 105 new
42 genomes) and we provide the first direct estimate of *T. gondii* mutation rate and the first
43 estimate of its generation time. We elucidate how the evolution of *T. gondii* populations is
44 intimately linked to the major events that have punctuated the recent history of cats. We
45 show that a unique haplotype —whose length represents only 0.16% of the whole *T. gondii*
46 genome— is common to all domestic *T. gondii* strains worldwide and has accompanied wild
47 cats (*Felis silvestris*) during their emergence from the wild to domestic settlements, their
48 dispersal in the Old World and their recent expansion to the Americas in the last six
49 centuries. By combining environmental and functional data to selection inference tools, we
50 show that selection of this domestic haplotype is most parsimoniously explained by its role
51 in initiation of sexual reproduction of *T. gondii* in domestic cats.

52 ## INTRODUCTION

53 *Toxoplasma gondii* is a zoonotic protozoan that has spread globally. This
54 apicomplexan parasite infects all warm-blooded species including humans, and its wide
55 range of host species suggests multiple routes for short and long-distance parasite
56 migrations (Galal et al., 2019). *T. gondii* is found in approximately 30% of the human
57 population and is the etiologic agent of toxoplasmosis, a disease causing a substantial public
58 health burden worldwide (Montoya and Liesenfeld, 2004). Infection with *T. gondii* has been
59 long considered as benign or even asymptomatic except for certain risk groups like the
60 developing foetus in case of congenital infection —with 200,000 new cases of congenital
61 toxoplasmosis each year (Torgerson and Mastroiacovo, 2013)— and immunocompromised
62 patients, for whom toxoplasmosis can have severe health consequences either during primo-
63 infection or reactivation. However, certain *T. gondii* populations have been associated with
64 severe toxoplasmosis in immunocompetent individuals (Carme et al., 2009; Pomares et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

65 2018; Schumacher et al., 2020). More importantly, an increasing number of epidemiological
66 studies suggest that chronic infection with *T. gondii* is associated with a wide variety of
67 neuropsychiatric disorders, substantially raising the public health importance of this global
68 and highly prevalent parasite (Milne et al., 2020). Given gaps in both the current preventive
69 (no vaccine available for humans) and therapeutic strategies (Innes et al., 2019;
70 Konstantinovic et al., 2019), active research to discover new ways to target this clinically
71 important protozoan are still needed.

72 *T. gondii* hosts get infected after ingestion of oocysts shed into the environment by
73 contaminated faeces of felids and develop persistent tissue-cysts. Another source of
74 infection for human and other meat-consuming species is the ingestion of raw or
75 undercooked meat from animals harbouring infective tissue-cysts. In the domestic
76 environment cats and rodents are considered as the most significant reservoirs for human
77 infection, since life cycle completion relies mainly on transmission between these two
78 categories of animal hosts, the rodents being the main prey of cats (Müller and Howard,
79 2016; Galal et al., 2020). Sexual recombination is possible when two different strains are
80 found simultaneously in the cat's gut. For this to occur a cat has to ingest within a few hours
81 two prey infected with different strains. There is therefore a time barrier for recombination
82 to occur, or alternatively the cat has to ingest a single prey infected with two different
83 strains (mixed infection), a rare event in nature given that intermediate hosts develop
84 immunity to new infections following their first infection.

85 From a genetic point of view, the population structure of *T. gondii* is characterized by
86 contrasting patterns of strain diversity mainly varying according to geographical origin and
87 ecotype. This diversity has mainly been identified based on the analysis of microsatellite
88 markers (MS), or Restriction Fragment Length Polymorphism (RFLP) (refer to Supplementary
89 Table 2 for correspondence MS and RFLP designations of lineages or genotypes). In the Old
90 World (Africa, Asia and Europe), most *T. gondii* isolates from humans, domestic animals and
91 wild fauna belong to few intercontinental clonal lineages: type I, type II, type III, Africa 1
92 (also designated as BrI) and Africa 4 (Shwab et al., 2014; Galal et al., 2019). Few other clonal
93 lineages have been described in certain countries such as Chinese 1 in China (Chaichan et al.,
94 2017) and Africa 3 in Gabon (Mercier et al., 2010), and strains not belonging to these major
95 lineages were rarely isolated (Galal et al., 2018). This genetic evidence argues that sexual

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

96 recombination between different strains is not frequent in these regions. In most regions of
97 the Old World, populations of wild felids have undergone massive decline (Goodrich et al.,
98 2015; Bauer et al., 2016), leaving domestic cats as virtually the only shedders of oocysts
99 capable of infecting domestic animals and humans, before spreading over long distances via
100 waterways to reach wildlife (Gotteland et al., 2014; VanWormer et al., 2013). In the New
101 World (North and South America), the most common Old World clonal lineages (type I, II, III
102 and Africa 1) are also found, in sympatry with a substantial diversity of local clonal lineages
103 strains and non-clonal strains specific to South or North America (Shwab et al., 2014; Jiang et
104 al., 2018). In contrast to the pattern observed in the Old World, the genotypic composition
105 of strains from wildlife differs importantly from strains commonly isolated in the domestic
106 environment (Mercier et al., 2011; Jiang et al., 2018). Globally, wild *T. gondii* populations —
107 isolated from wild animals or from humans in contact with wildlife and genetically distinct
108 from domestic *T. gondii* populations— are associated to biotopes where the presence of
109 wild felids is well-established (Carme et al., 2009; Khan et al., 2011; Mercier et al., 2011; L
110 Galal et al., 2019). This observation supports the notion that specific co-adaptations have
111 occurred between *T. gondii* strains and different feline species (Jewell et al., 1972; Miller et
112 al., 1972; Khan et al., 2014a).

113 To date, population genetic studies have only partially deciphered the phylogenetic
114 relationship between strains from different geographical areas. In particular, the
115 phylogenetic positioning of the most common clonal lineages relative to other *T. gondii*
116 lineages and populations remains unclear (Su et al., 2012; Lorenzi et al., 2016). In addition,
117 conflicting scenarios have been proposed to explain the global spread of these clonal
118 lineages (Minot et al., 2012; Bertranpetit et al., 2017; Shwab et al., 2018). Given the crucial
119 importance of domestic cats and rodents in the transmission of *T. gondii*, a presumed role of
120 these host species in the global spread of the major clonal lineages has been repeatedly
121 evoked in the literature (Lehmann et al., 2006; Shwab et al., 2018; Galal et al., 2019;
122 Hamidović et al., 2021). However, this hypothesis could not be formally tested previously
123 (Shwab et al., 2018; Hamidović et al., 2021) owing to the paucity of *T. gondii* samples in
124 many regions (Su et al., 2003; Lehmann et al., 2006; Khan et al., 2007; Minot et al., 2012;
125 Lorenzi et al., 2016). Moreover, lack of good estimates of parasite mutation rate and

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

126 generation time hampered attempts to date dispersal time in relation to expansion history
127 of principal hosts.

128 To address these questions, we generated the largest dataset of *T. gondii* genomes
129 produced to date (n = 156). We decipher the recent evolutionary history of *T. gondii*, using
130 both global and local ancestry analysis approaches. We provide the first direct estimate of
131 both *T. gondii* mutation rate and generation time that allowed dating major events shaping
132 parasite genome evolution. We uncover candidate genes whose geographic distributions,
133 genomic patterns and stage-specific patterns of expression are most parsimoniously
134 explained by local adaptation to the domestic ecotype and to transmission by domestic cats.

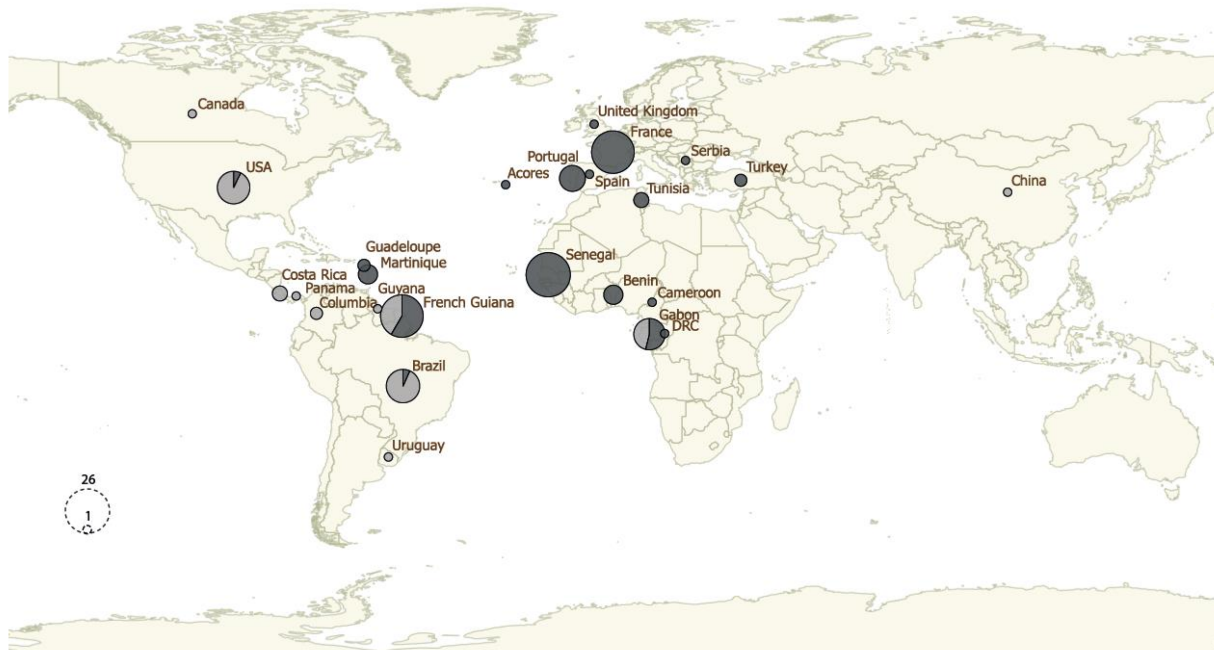
135

136 **RESULTS**

137 Paired-end sequence reads from 59 publicly available haploid genomes and from 105
138 new haploid genomes were aligned to the new PacBio reference assembly RH-88. These
139 samples included isolates from domestic (n=107) and wild animals (n=17), in addition to a
140 number of human isolates (n=42) (Supplementary Table 1). These samples originated from
141 22 countries (in addition to three French Overseas Departments), and covered most of the
142 global distribution of this species (Fig. 1). New identifiers were assigned to strains included in
143 this study, indicating the ecotype of the strain (Dc for domestic and Wd for wild) and the
144 country of origin. Note that most Old World samples included in this study belonged to
145 clonal lineages previously identified and defined from MS markers (types I,II, III, Africa 1 and
146 Africa 4; Supplementary Table 2), reflecting the very limited genetic diversity of *T. gondii* in
147 the Old World. Many of these samples originated from port regions in Africa (Goree island,
148 Saint-Louis, Dakar, Cotonou, Ouidah, Libreville) and Europe (Bordeaux, Le Havre), the most
149 likely source populations for recent human-mediated global expansion of *T. gondii* strains.
150 Isolates from the Americas originated from both coastal areas (e.g. Sao Paulo, Rio de Janeiro,
151 Cayenne) and inland areas, including wild environments in South and North Americas. In
152 addition, several isolates came from the Caribbean islands (Martinique, Guadeloupe), which
153 are well-known for their importance in maritime history linking Old and New Worlds.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

154 The 105 new genomes from this study were sequenced at a mean depth of 21X ,
155 ranging between 8 and 57X (Supplementary Table 1). In total, 156 genomes and 1,790,555
156 single-nucleotide polymorphisms (SNPs) passed all filtration criteria (see Methods). The
157 1,790,555 SNP-dataset was used for dating purposes. A second 1,262,582 SNP-dataset was
158 generated after removing singletons SNP from the first one and was used for population
159 genetics analyses.

160 

161 ***Fig. 1. The geographical distribution of* Toxoplasma gondii *strains analysed in this study.*** *Sizes of pie*
162 *charts correlate with total number of mono-strains isolates for each country. Isolates sequenced*
163 *specifically for this study are represented in dark grey and whole-genome sequence data from*
164 *previous studies —publicly available on the European Nucleotide Archive*
165 *(https://www.ebi.ac.uk/ena/browser/home)— are represented in light grey.*

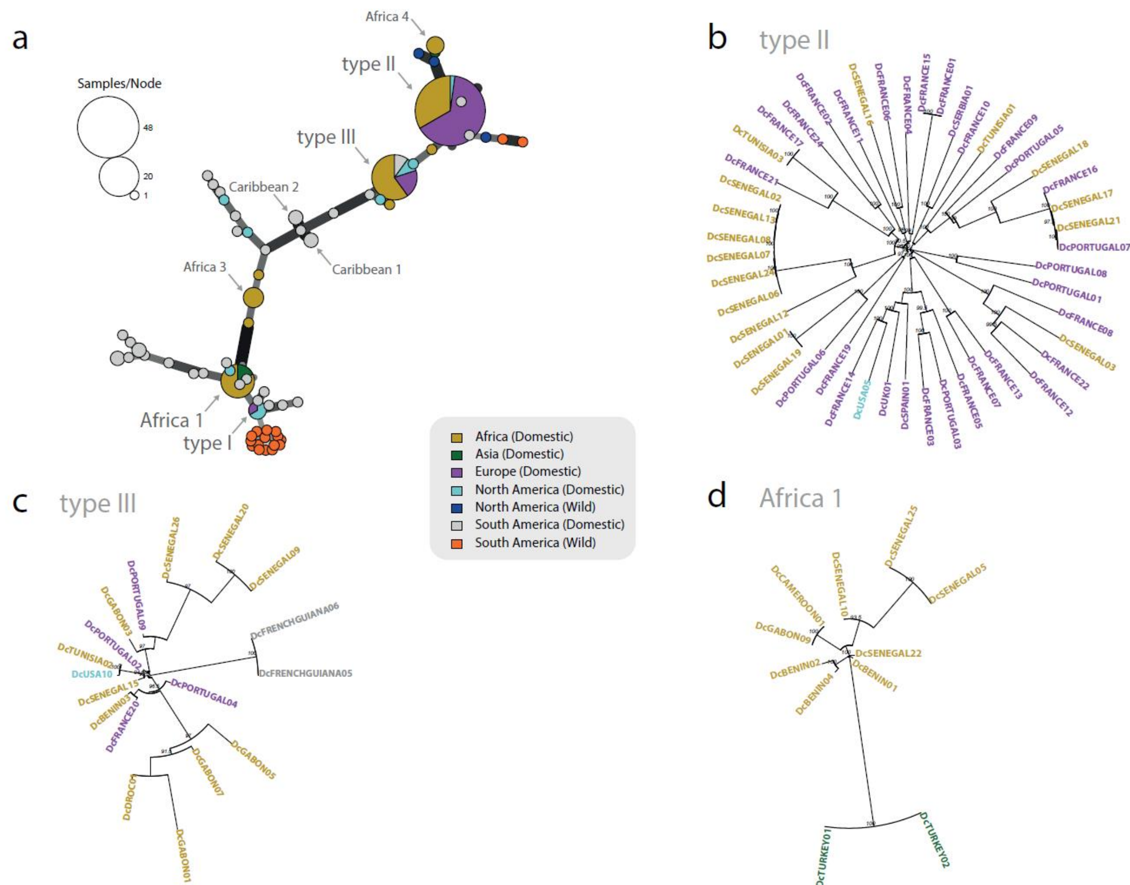
166

167 **Clonal lineages**

168 We first wondered whether the clonal lineages identified in *T. gondii* based on the
169 analyses of multilocus markers was evident at the genome scale. Our analyses (see
170 supplementary note for details) revealed the presence of four intercontinental clonal
171 lineages, as well as a number of regional clonal lineages (restricted to one continent) (Fig.
172 2a). By matching these results to previous findings relying on the analysis of MS markers, we
173 found that these four intercontinental clonal lineages correspond to type I (found in Asia,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

174 Europe, North and South America), type II (cosmopolitan), type III (cosmopolitan) and Africa
175 1 (found in Africa, South America and Western Asia) (Shwab et al., 2014; Chaichan et al.,
176 2017; Galal et al., 2018). Africa 4 is found in both Africa and Asia, although samples of this
177 lineage from Asia were not available for our study.



178

179 ***Fig. 2. Toxoplasma gondii clonal lineages description.*** *(a) Minimum spanning network of* T. gondii
180 *genomes. Genomes separated by a genetic distance less than or equal to 0.01 are collapsed in a*
181 *single circle and are considered belonging to the same clonal lineage. The size of each circle*
182 *corresponds to the number of individuals, and the colours indicate the continent of origin and the*
183 *ecotype of each individual. Thick and dark lines show MLGs that are more closely related to each*
184 *other whereas edge length is arbitrary. Neighbour-joining trees of* T. gondii *major lineages (b) type II,*
185 *(c) type III and (d) Africa 1. Individual genomes are colour-coded according to their continent of origin.*
186 *Support values greater than 90% using 1,000 bootstrap samples are shown.*

187

188 Within each clonal lineage, we did not find marked patterns of geographical structure
189 separating strains from the Old and New Worlds (Fig. 2 b-d). Conversly, strong clustering
190 between strains from different continents was repeatedly observed, suggesting recent

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

191 waves of intercontinental dissemination of these lineages. Four South American clonal
192 lineages could be identified although these lineages were undersampled. Two corresponded
193 to the previously described Caribbean 1 and Caribbean 2 lineages. Most domestic and wild
194 strains from South America were non-clonal.

195

196 **Global and local ancestry analyses.**

197 To decipher the genetic relationships between the *T. gondii* genomes from different
198 geographical origins, we first performed ancestry analysis using unsupervised clustering with
199 ADMIXTURE. This analysis was carried out after a step of clone-censoring of the dataset
200 (keeping only one representative strain of each clonal lineage / type), resulting in a dataset
201 of 71 strains and 588,777 SNPs. The optimal number of ancestral groups was determined to
202 be five (lowest cross-validation error), but we also examined different K values. Old World
203 strains were mostly a mixture of different intercontinental clonal lineages (types I, II, III,
204 Africa 1 and Africa 4) (Supplementary Fig. 1). In the New World, the Wild Amazonian group
205 (in orange) constituted a well-defined non-admixed ancestral group, clearly divergent from
206 the different intercontinental clonal lineages at different K solutions. In addition, a wild
207 group of strains from the Amazonian forest in South America and of a wild strain from North
208 America could be distinguished at K=9, and was designated as Pan-American. Other New
209 World lineages and strains were composed of New World-specific ancestries (in burgundy
210 and in orange) and intercontinental lineages ancestries, with many of these strains exhibiting
211 a mixed pattern of these two categories of ancestry in the same genome. This latter pattern
212 was mainly noticed among domestic strains, and could be suggestive of hybridization events
213 between intercontinental lineages and New World specific clades.

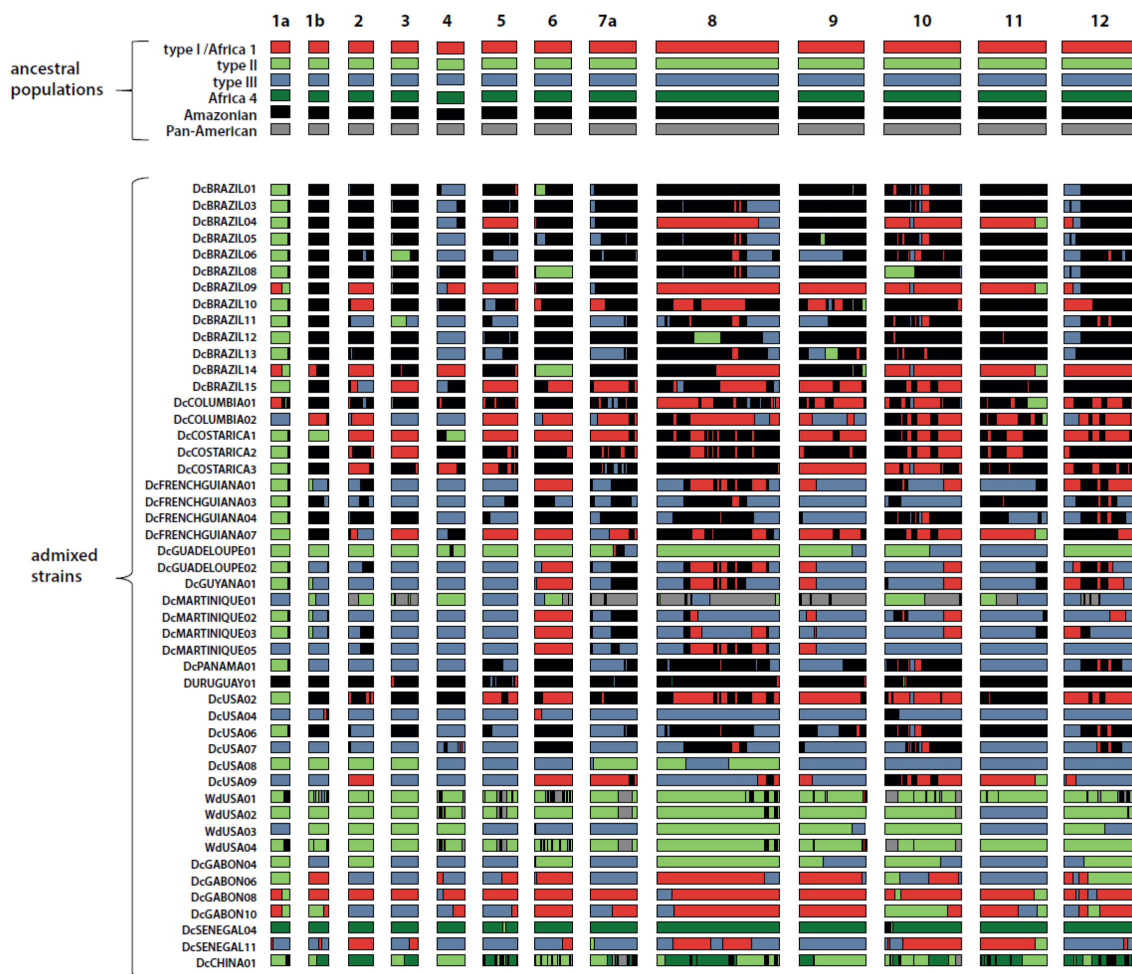
214 To better understand the patterns of admixture, we generated a co-ancestry matrix
215 with whole nuclear genome data and independent co-ancestry matrices for each of the 13
216 nuclear chromosomes using ChromoPainter. This revealed that domestic strains from the
217 New World all shared chunks of chromosomes with at least one of four intercontinental
218 domestic clonal lineages type I, type II, type III and Africa 1 (Supplementary Fig. 2 a-n) and
219 for many of them the wild American strains (Pan-American or Amazonian). By contrast, the
220 four intercontinental domestic clonal lineages did not share any chromosome regions with

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

221 the two above mentioned wild American populations. Accordingly, we performed local
222 ancestry analyses by defining the intercontinental lineages (type I, type II, type III, Africa 1
223 and Africa 4) and the unadmixed wild populations (Amazonian and Pan-American) as
224 putative ancestral populations. Putative hybrids (all the other genomes) were made up of
225 few large blocks of different ancestries (up to five ancestors for the same strain), and whole
226 chromosomes of single ancestry are also often observed (Fig. 3). At least one large segment
227 (> 1Mb) of ancestry corresponding to an intercontinental lineage was found in all domestic
228 strains from both Old and New Worlds. One exception was DcURUGUAY01 (CASTELLS) that
229 had almost only an Amazonian ancestry. Amazonian ancestry was identified in nearly all
230 domestic strains from South and Central America, and to a lesser extent in North America,
231 but was absent from almost all Old World domestic strains. Pan-American ancestry was
232 much rarer among putative hybrids, although it was found in both South and North America,
233 including the wild putative hybrid strains from North America. Analyses of the apicoplast
234 genome (a maternally inherited organelle that does not undergo recombination) confirmed
235 the hybrid origin of the American domestic strains, as each apicoplast genome harboured a
236 single ancestry, related either to an intercontinental lineage or to a wild New World
237 population (mainly Amazonian) (Supplementary Fig. 4). In the Old Word, most admixed
238 strains were the result of an admixture between the major lineages, but these hybrids
239 remain rare compared to the American continent where almost all domestic strains have a
240 hybrid origin.

241 Overall, our results demonstrate that Old and New World *T. gondii* parasites present
242 radically different patterns of genetic diversity. In the Old World (Europe, Africa and Asia),
243 most strains belong to one of the previously defined intercontinental clonal lineages with
244 rare hybrids observed between the different lineages. On the other hand, most New world
245 parasites isolated from wild animals form well-defined non-clonal genetic clusters and those
246 isolated from the domestic environment are the result of hybridizations between
247 intercontinental lineages and a number of New World-specific wild populations
248 (Supplementary Fig. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



249

250 ***Fig. 3. Genome-wide distribution of ancestry in all putative hybrid* Toxoplasma gondii *genomes.***
251 *Plots are graphically displayed using karyoploteR (Gel and Serra, 2017) and show ancestry estimates*
252 *at each genomic position for the 13 nuclear chromosomes. Colours reflect putative ancestral*
253 *populations.*

254

255 Timing of emergence, spread and introgressions of intercontinental *T. gondii* clonal 256 lineages.

257 Using stored aliquots of two long-term *in-vivo* cultured *T. gondii*, we estimated the
258 mutation rate of the parasite (see Supplementary Methods). This was complemented with
259 data from the literature to estimate the parasite generation time. For each clonal lineage
260 using these estimates (ranging from 3.1×10^{-9} to 11.7×10^{-9} mutations per site per year) we
261 first determined the time to the most recent common ancestor (TMRCA) between the two
262 most divergent genomes (Fig. 4a). Regarding the four major intercontinental clonal lineages,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

263 our estimates showed that type II emerged 12,980-48,988 years ago, whereas type I, type III
264 and Africa 1 emerged much later: 651-2457, 683-2,578 and 629-2,376 years ago,
265 respectively. Note that inclusion of additional isolates from the same or different geographic
266 areas could alter these estimates by revealing more ancient divergence times between
267 strains of each respective lineage. This is particularly true for type I for which only three
268 samples were available. Regarding the four South American-specific clonal lineages, the
269 divergence times estimated between the most distant isolates ranged between 197 -742 and
270 225 - 850 years ago, for Caribbean 1 and Caribbean 2, respectively.

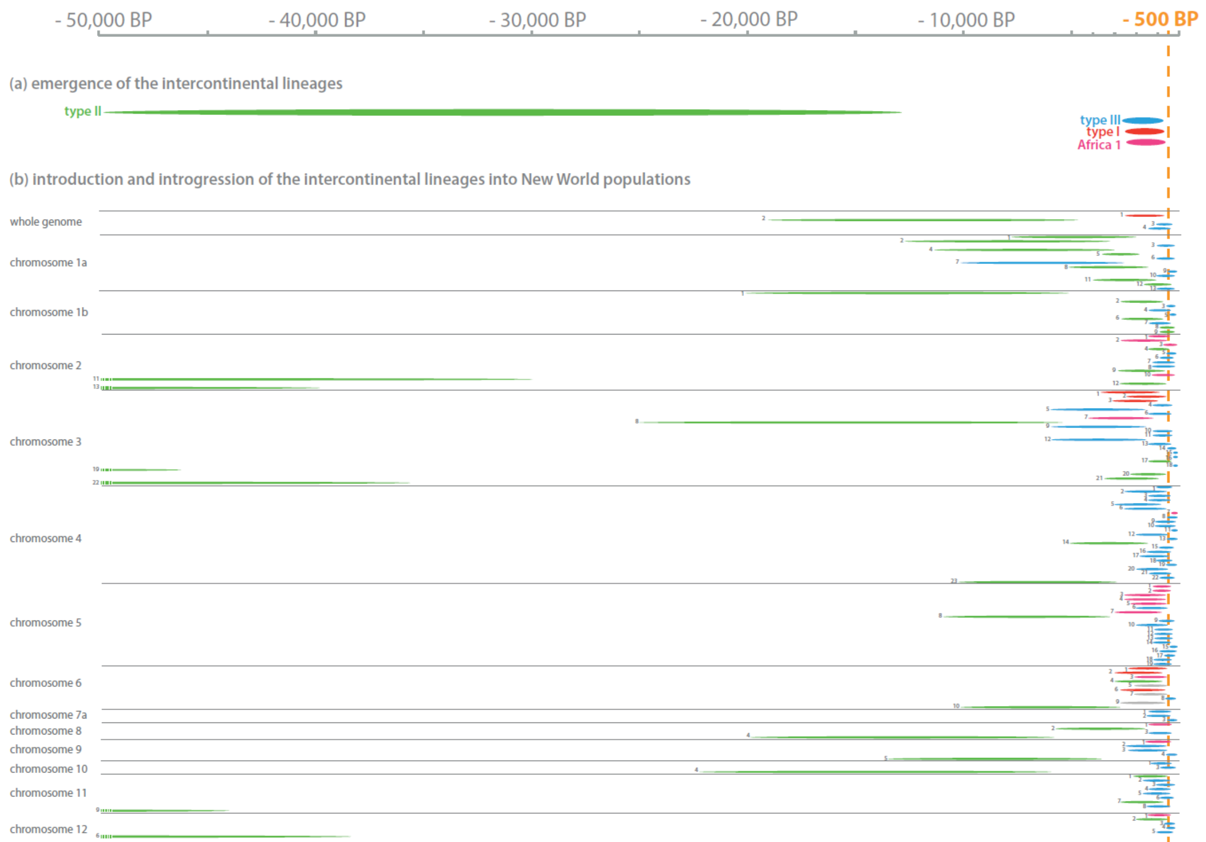
271 We next explored the divergence time between New and Old World strains sharing
272 the same ancestry. We calculated TMRCA between New World strains and their closest
273 relative from the Old World belonging to the same intercontinental clonal lineage (Fig. 4;
274 Supplementary Table 3). In addition, we calculated TMRCA between full chromosomes of
275 New World hybrid strains inherited from one of the four intercontinental lineages and their
276 closest relative among Old World strains having the same respective chromosomal ancestry
277 (Fig. 4; Supplementary Table 4). Note that TMRCA estimates are also very sensitive to
278 sampling and obtaining accurate estimates depends on robust sampling of source
279 populations (populations that expanded to the New World). Among isolates of each
280 respective clonal lineage from Africa and Europe, isolates not belonging to source
281 populations are expected to exhibit some degree of divergence from these source
282 populations, and consequently from individuals that have expanded to the New World from
283 these source populations. This divergence bias is expected to be more substantial in more
284 diversified lineages (type II *versus* other intercontinental lineages) owing to greater
285 divergences between strains of the same lineage, even within a single region (also refer to
286 Fig. 2b). In this sense optimal sampling would have included isolates from all port areas
287 historically involved in transatlantic and colonial trade in Europe and Africa, which is only
288 partially the case in this study.

289 Most estimates of TMRCA indicate that massive introgressions of types I, II, III and
290 Africa 1 into New World populations have occurred in the last few centuries, and provide
291 evidence for very recent migrations of strains between Old and New Worlds (Fig. 4b). A
292 number of TMRCA estimates (in particular among strains having type II ancestry) indicated

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

293 older divergence times, which could be explained by older migratory waves, or by a
294 divergence bias related to sampling.

295



296

297 ***Fig. 4. Dating estimates for emergence of intercontinental lineages and their expansion to the New***
298 ***World.*** *(a) The length of coloured ellipses represents the estimated interval for the time to the most*
299 *recent common ancestor (TMRCA) between the two most divergent genomes of each intercontinental*
300 *lineage: type I, type II, type III and Africa 1. (b) The length of coloured ellipses represents the*
301 *estimated interval for the TMRCA between full chromosomes of New World strains and their closest*
302 *relative from the Old World having the same respective chromosomal ancestry: type I in red, type II in*
303 *green, type III in blue, Africa 1 in purple and Africa 3 in grey. Precise TMRCA estimates are presented*
304 *in Supplementary Table 3 and 4. A numeral is attributed to each TMRCA estimate on the figure to*
305 *facilitate the correspondence with the Supplementary Table 3 and 4.*

306

307 Conversely, the wild strains of type 12 (WdUSA01 and WdUSA04) and type II, although
308 having chromosomes of the same ancestry, had TMRCA estimates consistent with a
309 divergence prior to the emergence of type II lineage (Fig. 4; Supplementary Table 4). It

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

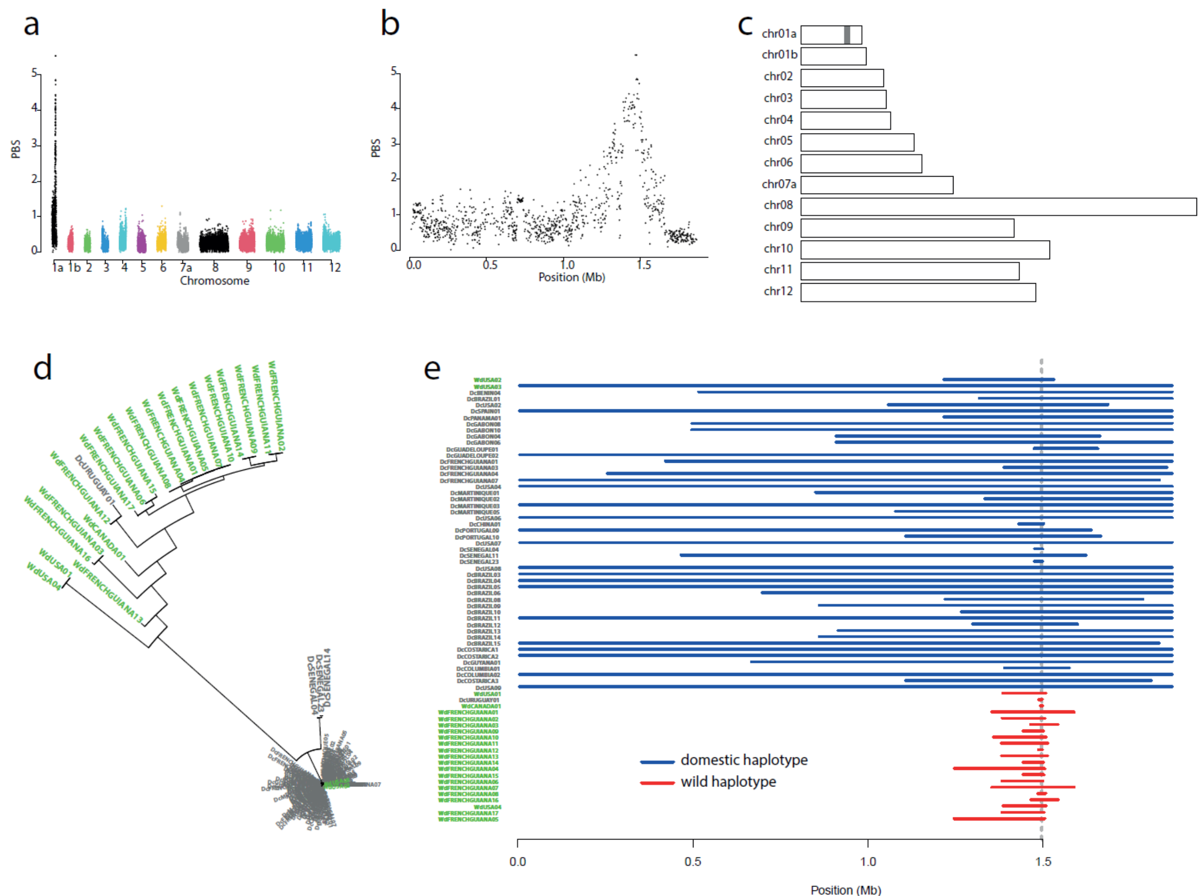
310 confirms that type 12 is a true wild population, sharing no recent ancestry with the domestic
311 type II lineage. WdUSA02 (B41) showed an admixed pattern of divergence from the type II
312 lineage, both recent and more ancient, consistent with a recombination between a domestic
313 type II and a wild type 12 strain. WdUSA03 (B73), although isolated in the wild (from a bear)
314 had a chromosomal ancestry consistent with recent recombination having occurred between
315 type II and type III lineages. This latter observation is consistent with evidence from
316 multilocus markers of a dissemination of domestic strains into the wild ecotype in North
317 America (Jiang et al., 2018).

318

319 **Candidate genes contributing to adaptation to the domestic environment.**

320 We asked whether hybrid domestic strains from the New World are the result of
321 random recombinations between intercontinental domestic lineages and New World-
322 specific —wild— clades, or whether certain domestic alleles inherited from intercontinental
323 lineages had been selected during this process. This could explain the greater success of
324 hybrid strains compared to wild strains in occupying domestic niches in New World
325 countries. To do so, the 71 genomes of the clone-censored dataset were split into three
326 groups based on results from global and local ancestry analyses: (1) domestic populations
327 (intercontinental lineages and hybrid strains having a recent shared ancestry with these
328 lineages), (2) wild populations (Amazonian, Pan-American and type 12), and (3)
329 DcURUGUAY01 (CASTELLS). This latter strain was used as outgroup given its New World-
330 specific ancestry (Fig. 3), and its divergence from other populations (Lorenzi et al., 2016). The
331 wild strains WdUSA02 and WdUSA03 were included in the domestic group given their
332 recently inherited large genomic segments from domestic lineages of types II and III (Fig. 3
333 and 4; Supplementary Table 4). We computed the population branch statistic (PBS) using
334 PBScan in order to detect genomic regions of unexpectedly high divergence between the
335 three groups. A unique outlier region of marked divergence between input groups was
336 identified (Fig. 5a); it occurred on chromosome 1a over a region ~600 kb long (Fig. 5b). In
337 parallel, we scanned the nuclear genome for “domestic” variants (alleles common to
338 domestic strains and not found in wild strains). We chose to include DcURUGUAY01 in the
339 wild group given its nearly exclusive Amazonian ancestry (Fig. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



340

341 ***Fig. 5. Outlier selection region for adaptation of* Toxoplasma gondii *to the domestic ecotype.*** *(a)*
342 *The Manhattan plot shows the genome-wide distribution of population branch statistic (PBS)*
343 *calculated between wild and domestic strains for each 50-SNP sliding window using the clone-*
344 *censored dataset (n=71). (b) This is a zoom on PBS obtained for chromosome 1a. (c) Genomic*
345 *positions of domestic variants (n=310) from the clone-censored dataset are shown in grey. (d)*
346 *Neighbour-joining tree of the outlier selection region (positions 1,390,579 to 1,502,589 on*
347 *chromosome 1a). It was produced using ape R package by computing genetic distances based on the*
348 *dissimilarity matrix produced by poppr R package. It includes all the strains of the dataset (n=156).*
349 *Individual strains are colour-coded according to ecotype of origin (domestic strains in grey and wild*
350 *strains in green). (e) Haplotype length around the position 1495970 on chromosome 1a in domestic*
351 *strains relative to wild strains of the clone-censored dataset. The plot shows the boundaries of the*
352 *longest shared haplotype (the range over which it is identical to at least one other haplotype) around*
353 *the domestic allele of the focal marker (chr01a_1495970) relative to the wild allele.*

354

355 In total, 310 variants were identified, of which 58 were missense variants (Supplementary
356 Table 5). Strikingly, all 310 variants occurred within the outlier region identified using PBS
357 statistic, over a region ~150 kb long containing 25 genes (Fig. 5c), between positions
358 1,352,873 and 1,501,765 (Supplementary Table 6). When we consider the ancestry pattern

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

359 of this region (Fig. 3), we notice that nearly all hybrids strains have a type II/type III ancestry
360 between positions 1,390,579 and 1,502,589. This latter genomic region of ~100 kb
361 corresponds to a remarkably conserved haplotype shared by all domestic strains and clearly
362 divergent from wild haplotypes (Fig. 5d). Interestingly, only Africa 4 strains showed some
363 degree of divergence from other domestic strains, although they carried all 310 domestic
364 variants identified within the outlier region of selection. Extended linkage disequilibrium was
365 observed around this outlier region in domestic strains relative to wild ones, a signal of
366 positive selection acting specifically on domestic strains (Fig. 5e).

367 In order to determine which gene(s) is/are most likely associated to adaptation to the
368 domestic ecotype, we analysed the function of the 25 candidate genes present in the ~150
369 kb genomic region (carrying the 310 domestic variants) and their stage-specific pattern of
370 expression. Although 14 genes were annotated for protein function, data was lacking in the
371 literature about possible associations between allelic heterogeneity and differential
372 adaptations to specific hosts or ecotypes. By focusing on stage-specific patterns of
373 expression, we found that six genes displayed stage-specific expression as they were
374 expressed at only one stage (Ramakrishnan et al., 2019; Farhat et al., 2020). Among them
375 two genes were found to be only expressed during enteroepithelial stages (EES) of
376 development (early cellular forms characteristic of the onset of the sexual stage in cat
377 enterocytes): TGRH88_020260 (TGME49_295995) and TGRH88_020330 (TGME49_295920).
378 Missense variants segregating domestic from wild strains were only found for
379 TGRH88_020330 (TGME49_295920). This gene is annotated as encoding a hypothetical
380 protein. Five missense variants segregating domestic strains from wild strains were found for
381 this gene (Supplementary Tables 5 & 7).

382

383 **DISCUSSION**

384 **Intercontinental lineages spread from the Old to the New World.**

385 The global dataset of whole genome sequences studied herein confirm previous
386 findings based on multilocus genotyping (mainly MS

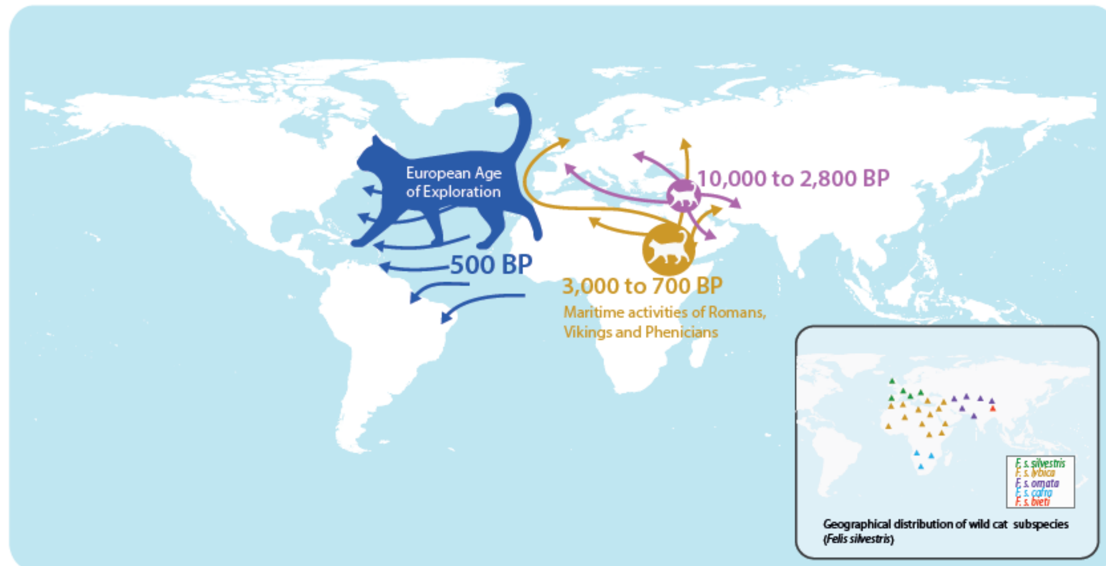
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

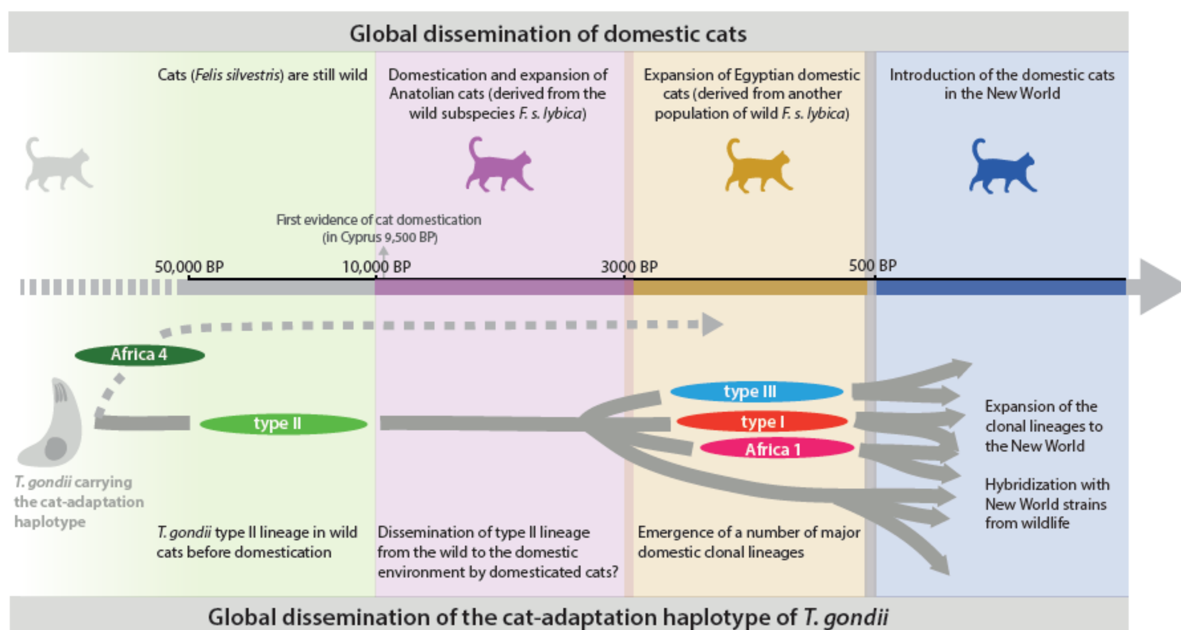
387 and RFLP) of a strong clonal structure for most *T. gondii* populations (Su et al., 2012; Galal et
388 al., 2019). We provide strong evidence of intercontinental dissemination of the most
389 prevalent lineages (types I, II, III and Africa 1) using a strictly clonal mode of propagation. All
390 these four lineages are well adapted to transmission by the domestic cat as they are
391 frequently isolated in the domestic environment (Chaichan et al., 2017; Galal et al., 2018;
392 Shwab et al., 2018).

393 We estimated that *T. gondii* lineage type II emerged 12,980-48,988 years ago (Fig. 6).
394 At this period, cat domestication was still an ongoing process, and it is therefore likely that
395 type II lineage was circulating in wild cats (*Felis silvestris*). The domestication of cats first
396 occurred in the Near East, coincident with agricultural village development in the Fertile
397 Crescent about 10,000 years ago (Driscoll et al., 2007; Ottoni et al., 2017). This *T. gondii*
398 lineage could have accompanied cats of subspecies *F. s. lybica* (one of the five known clades
399 of wild cats of species *Felis silvestris*, found in different regions of the Old World) that
400 established themselves in human settlements in the Near East, or emerged later in the
401 domestic environment from other wild cat populations in Africa, Asia or Europe following
402 the expansion of the Neolithic revolution to those regions. Type I, type III and Africa 1
403 lineages emerged much later, 651-2457 years, 683-2,578 years and 629-2,376 years ago,
404 respectively. Importantly, these domestic lineages emerged before dissemination of
405 domestic cats to the New World 500 years ago; it is hence likely that they emerged in the
406 Old World. From the beginning of the Iron age, about 3,000 years ago, maritime
407 transportation substantially increased in the Old World (Jones et al., 2013). Maritime
408 activities of Romans in Antiquity (Peters, 1998), of Phoenician during medieval times
409 (Bonhomme et al., 2011) and of Vikings between the 7th and the 11th centuries (Jones et al.,
410 2012), contributed to the expansion of domestic cats, mice and rats throughout the
411 Mediterranean basin and to Continental Europe. During these periods, Egyptian domestic
412 cats gradually took over other populations of domestic cats such as in Anatolia before
413 spreading to most areas of the Old World (Ottoni et al., 2017). These movements of *T. gondii*
414 hosts could have promoted encounters with previously allopatric populations of *T. gondii*
415 (e.g. between Europe and Africa), fostering the emergence of new lineages by hybridization.
416 This hypothesis is supported by the shared ancestry that is observed on certain genomic
417 regions between the most common lineages. Boyle et al. (2006) showed that types I and III

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

418 are respectively second- and first-generation offspring of a cross between a type II strain and
419 one of two unknown ancestral strains. The most parsimonious scenario is that these lineages
420 emerged following the expansion of the geographical range of type II strains in the Old
421 World during this period.





422
423 ***Fig. 6. Graphical summary of major events that have punctuated the recent evolutionary history of***
424 **Toxoplasma gondii *in relation to history of cats' dispersal.*** *Geographical distribution of wild cat*
425 *subspecies and historical data about cats' dispersal are derived from Ottoni et al. (2017). The grey*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

426 *dotted arrow in the bottom part of the figure indicates that emergence of Africa 4 lineage could not*
427 *be dated in this study due to a lack of samples belonging to this lineage.*

428

429 **Introgression of the Old World domestic lineages into New World populations.**

430 A previous study showed that common inheritance of large haploblocks is the major
431 factor in determining the phylogenetic grouping of *T. gondii* strains (Lorenzi et al. 2016).
432 Here, we demonstrate that nearly all New World domestic strains are hybrid strains
433 harbouring large chromosomic regions of either types I, II, III or Africa 1 ancestry. Moreover,
434 many TMRCA estimates suggest that hybridization events that gave rise to these strains
435 occurred after the introduction of these lineages in the Americas in the last six centuries.
436 These estimates coincide to the onset of the European "age of exploration" (Subrahmanyam
437 and Alam, 2007). During this period, human activities enabled domestic cats, but also mice
438 (*Mus musculus*) and rats (*Rattus rattus* and *Rattus norvegicus*) to reach the Americas for the
439 first time (Lipinski et al., 2008; Macholán et al., 2012; Puckett et al., 2016), allowing for the
440 first time the emergence of a domestic cycle of *T. gondii* in those areas.

441 It is noteworthy that hybrid strains are much more common in the New World
442 (especially in South America) compared to the Old World indicating —currently or in the
443 past— more frequent recombination events. Most hybrid strains appear to be the results of
444 only one or a few rounds of meiotic reproduction when considering the chromosomal
445 pattern of ancestry of experimental hybrids (Khan et al., 2014b). Indeed, we did not observe
446 a fine mosaic of different ancestries alternating across genomes, as is usually observed when
447 sexual recombination often occurs in a population (Henn et al., 2012; Fitak et al., 2018; Kim
448 et al., 2020). Sexual recombination in *T. gondii* is favoured by mixed infections in cat prey,
449 which is limited by the immunity developed by an intermediate host following its primary
450 infection (referred to in the Introduction). This immunity often protects the intermediate
451 host from new infections with different strains, but not from highly divergent strains as
452 found in South America (Elbez-Rubinstein et al., 2009). We propose that following their
453 introduction in the Americas, rodents infected with Old World lineages were exposed to
454 highly divergent strains from the wild environment near human settlements. Their immunity
455 being unable to contain these new infections, the rodents could have become superinfected

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

456 with highly divergent strains giving rise to tissue-cysts of New World strains alongside tissue-
457 cysts emanating from their primary infection. This unique situation would provide
458 favourable conditions for the emergence of hybrid populations due to cats feeding on these
459 superinfected intermediate hosts. The emergence of big cities, the decline of wildlife and the
460 great proliferation of domestic cats would have gradually limited the exposure of domestic
461 intermediate hosts to wild strains in South and North America.

462 We show that for certain chromosomes type II has sister clades (close but distinct)
463 among wild strains from North America. The time to the most recent common ancestor
464 (TMRCA) estimated for domestic type II and wild type 12 (WdUSA01 and WdUSA04) on
465 several chromosomes is clearly anterior to the onset of domestication and to the emergence
466 of type II. Evidence from apicoplast sequences also shows that Asian Chinese 1 shares a
467 common ancestor with wild type 12. Note that all these strains belong to the same clade
468 (refer to Supplementary Fig. 1). These data support the occurrence of *T. gondii* migrations
469 between Asia and North America, probably anterior to the Neolithic revolution and the
470 domestication era. Migrations were probably mediated by movements of animal herds
471 through the land bridge formed by the Bering Strait during the late Pleistocene period until
472 about 13,000 years ago. During this period a corridor was created by falling sea levels that
473 provided an opportunity for Asian species including mammoths, bison, muskoxen, caribou,
474 lions, brown bears, and wolves to move into North America (Guthrie, 2004; Lowe and
475 Walker, 2014; Froese et al., 2017; Phillips et al., 2018). Assuming a role of late Pleistocene
476 animal species in disseminating *T. gondii*, an Asian origin of this clade appears more likely
477 given the direction of migrations inferred for these animals. It is consistent with the
478 hypothesis of an Old World origin of type II lineage, as previously suggested by Shwab et al.
479 (2018) using multilocus markers and not a North American origin as proposed by other
480 studies (Khan et al., 2007; Minot et al., 2012).

481

482 **A candidate gene for adaptation of *T. gondii* to the domestic cat.**

483 We sought to identify genes under positive selection for adaptation of *T. gondii*
484 strains to the domestic ecotype. Our genome-wide scan for selection identified a unique
485 genomic region of ~100 kb (0.16% of the whole *T. gondii* genome) on chromosome 1a

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

486 exhibiting a nearly perfect dichotomy between wild strains and domestic strains from all
487 over the world. The few experimental infections carried out on domestic cats showed that
488 cats infected with domestic lineages (carrying this domestic haplotype) produce oocysts
489 more efficiently compared to when they are infected with wild strains (Khan et al., 2014b).
490 The signal we found of a unique global haplotype common to domestic strains is therefore
491 probably the results of an adaptation to the domestic cat (*Felis catus*), not forgetting that
492 the latter is the only indispensable host species for the transmission of *T. gondii* in the
493 domestic environment.

494 Note that a number of strains shared the same haplotype for the whole length of
495 chromosome 1a, a pattern previously noticed in past studies (Khan et al., 2007; Lorenzi et
496 al., 2016). We provide strong evidence that this pattern is specific to domestic strains. It can
497 be explained by the much stronger linkage disequilibrium observed in domestic strains
498 relative to wild ones around the outlier region identified in this study due to selection, often
499 reaching the whole length of chromosome 1a. Given the rarity of sexual recombination in *T.*
500 *gondii* populations, it is likely that a number of domestic lineages and strains inherited the
501 entire chromosome 1a from an ancestor carrying the advantageous allele. Other domestic
502 strains inherited a more or less important portion of chromosome 1a from this ancestor,
503 leading to the tightening of the inherited portion around the domestic haplotype under
504 selection following successive rounds of sexual recombination.

505 This domestic haplotype exhibited some degree of genetic divergence between Africa
506 4 strains and all other domestic strains. This divergence probably occurred in the Old World
507 given the exclusive occurrence of Africa 4 lineage in Africa and Asia. This observation
508 provides additional evidence for an Old World origin of the "cat adaptation haplotype", not
509 forgetting that it is also where domestic cats first emerged. It is therefore likely that this
510 haplotype spread to the domestic environment from two different sources: the Africa 4
511 lineage that expanded in Africa and Asia, and the type II lineage (the presumably oldest
512 domestic lineage) at the origin of the global expansion of the haplotype.

513 Within the outlier region of selection, TGRH88_020330 (TGME49_295920) was
514 selected as the top candidate gene for adaptation to domesticity, as it was the only gene
515 specifically expressed by *T. gondii* during its early stages of sexual reproduction in cat

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

516 enterocytes that also carried functionally relevant (missense) variants (n = 5) segregating
517 domestic *T. gondii* strains from wild ones. However, it was not possible to determine which
518 of these missense variants has functional relevance. The genomic region carrying this gene
519 exhibits strong linkage disequilibrium among domestic strains implying that certain variants
520 have been fixed by hitch-hiking. The expression of this gene of unknown function begins at
521 the earliest stages of sexual multiplication (Ramakrishnan et al., 2019; Farhat et al., 2020),
522 suggesting that it could have a role in specific host recognition, to enable the initiation of
523 sexual reproduction when a given *T. gondii* strain infects the proper host species within the
524 felidae family. The highly successful haplotype carried by this genomic region and shared by
525 almost all domestic strains probably enables efficient parasite transmission by present-day
526 domestic cats, not forgetting that according to our dating estimates all variants constituting
527 this haplotype were already fixed before the onset of domestication. It suggests that wild
528 strains harbouring this haplotype before domestication must have been efficiently
529 transmitted by wild ancestors of the present-day domestic cats.

530 Di Genova et al. (2019) have recently produced cat intestinal organoids for
531 experimental purposes, an important breakthrough to study sexual reproduction of *T. gondii*
532 knowing the important ethical concerns associated with the use of live cats. Developing
533 similar experimental models from intestinal cells of wild felids could enable a more accurate
534 understanding of the function of the top candidate gene and other candidate genes
535 occurring within the outlier region of selection.

536 In summary, we have produced a large dataset of high-quality *T. gondii* genomes and
537 estimated the parasite's mutation rate and generation time. We dated the emergence of the
538 most common *T. gondii* clonal lineages, their recent dispersal and introgressions into New
539 World populations of *T. gondii*. We show that the substantial diversity of domestic strains
540 found in the New World is the result of hybridizations between four recently introduced Old
541 World domestic lineages (adapted to domestic cats) and New World strains from wildlife. A
542 unique cat-adaptation *T. gondii* haplotype —today carried by almost all domestic strains
543 worldwide— has been largely conserved since its initial emergence from wilderness to
544 domestic settlements, and during its dissemination in the Old World and its recent
545 expansion to the New World. The selection of this now global domestic *T. gondii* haplotype is
546 most parsimoniously explained by its role in the initiation of sexual reproduction of *T. gondii*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

547 in domestic cats. Importantly, parasite gene(s) involved in the initiation of sexual
548 reproduction could be promising targets for the development of a cat vaccine. In the context
549 of a One Health integrated vaccine programme (Innes et al., 2019), controlling oocysts
550 excretion by domestic cats is considered crucial, since it could be the most efficient way to
551 break the cycle of transmission, limit environmental contamination by oocysts and prevent
552 infection of other T. gondii hosts including humans.

553

554 METHODS

555 We studied 106 *T. gondii* isolates provided by the French Biological Resource Centre
556 (BRC) for *Toxoplasma* (http://www.toxocrb.com/). This certified structure (NF S96-900
557 standard) manages the storage of T*. gondii* strains from human or animal toxoplasmosis to
558 make them available to the scientific community. Our analyses were complemented with
559 whole-genome sequence data of 59 strains (100 bp paired-end reads) from a previous study
560 (Lorenzi et al., 2016) made publicly available on the European Nucleotide Archive
561 (https://www.ebi.ac.uk/ena/browser/home).

562 Following parasite culture (Supplementary Methods), total genomic DNA was
563 extracted from 200µl of tachyzoites suspension, using the QIAamp DNA MiniKit (Qiagen,
564 Courtaboeuf, France). *Toxoplasma gondii* DNA extracts were genotyped using 15 MS
565 markers in a single multiplex PCR-assay, as described previously (Ajzenberg et al., 2010).This
566 step was necessary to check for cross-contaminations between samples during culture and
567 to identify mixed infections. A mixed infection was identified in one isolate (FR-Mac fas-002;
568 *Macaca fascicularis* ; Mauritius) by the presence of two alleles at 12 loci; only one strain was
569 genotyped at the time of isolation, and it is therefore likely that the second strain initially
570 had a lower tissue load in the infected host, and took longer to grow to levels sufficient for
571 library construction. Only the mono-strain isolates (n=105) were sequenced. DNA was
572 sheared into 400–600-base pair fragments by focused ultrasonication (Covaris Adaptive
573 Focused Acoustics technology, AFA Inc, Woburn, USA). Standard indexed Illumina libraries
574 were prepared using the NEBNext DNA Library Prep kit (New England BioLabs), followed by
575 amplification using KAPA HiFI DNA polymerase (KAPA Biosystems). 150 bp paired-end reads
576 were generated on the Illumina NextSeq 500 according to the manufacturer's standard

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

577 sequencing protocol. The 105 samples sequenced for this study are deposited in ENA under
578 the study accession number XXXXXXXXX.

579 **Mapping and variant calling.**

580 Recent advances in variant calling enabled the use of publicly available panels of
581 validated single-nucleotide polymorphisms (SNPs) and indels to estimate the accuracy of
582 each base call and minimize the generation of false positive SNPs. Unfortunately, building
583 these panels is labour-intensive and such data is lacking for under-studied organisms such as
584 *T. gondii*, which still falls under the category of "non-model" organisms. Ribeiro et al. (2015)
585 explored the relationship between the choice of tools and parameters in non-model
586 organisms, their impact on false positive variants, and formulated recommendations for
587 variant calling. Here, we followed their recommendations in order to minimize the call of
588 false positive variants, which was a critical point regarding our objective of estimating the
589 occurrence times of recent events in *T. gondii* evolution. In this sense, reads were submitted
590 to a stringent mapping configuration (not more than 2% of mismatches), by using BWA 0.7
591 against the newly available PacBio reference genome RH-88 (13 nuclear chromosomes that
592 cover 63.97 Mb; available on https://toxodb.org; release date 2020-05-15). Mapped reads
593 were sorted with Samtools 1.11 and duplicate reads were marked with Picard
594 'MarkDuplicates' 2.25. Individual BAMs were subsequently merged, and variant calling was
595 performed with FreeBayes 1.3.5 that is considered better than the routinely used Genome
596 Analysis Toolkit (GATK) in non-model organisms (Ribeiro et al., 2015; Calarco et al., 2018). At
597 the individual level, alignments having a mapping quality (--min-mapping-quality) less than
598 20, a coverage (--min-coverage) less than 3, and alleles having a supporting base quality (--
599 min-base-quality) less than 20 were excluded from the analysis. Genotype calls having
600 fraction of conflicting base calls of more than 10% were also excluded. Finally, only
601 individuals having missing genotype data less than 5% were kept for subsequent analyses to
602 minimize false negative calls. At the population-level, SNPs having a high missing genotype
603 rate (>10%) were filtered out. With the above filters in place, two individuals were excluded
604 (missing genotype data >5%), before excluding an additional 6 samples due to unreliable
605 information about the country of origin and/or the ecotype. The filtered set of SNPs was
606 annotated with snpEff 5.0 6 using the RH-88 annotation file.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

607 For the 35kb *T. gondii* apicoplast genome, mapping of reads was performed against
608 the ME49 reference genome assembly (release date 2013-04-23), as the PacBio reference
609 genome RH-88 had a high proportion of low complexity sequence (~70% in RH-88 genome
610 *versus* ~30% in ME49 reference genome; data not shown). The mapping quality of reads in
611 these low complexity regions was too low and therefore these regions were not exploitable
612 for sequence analysis. A mapping configuration and variant calling parameters identical to
613 those used for the nuclear genome were used for the apicoplast genome. In addition to the
614 six samples having unreliable information about the country of origin and/or the ecotype, six
615 additional samples were excluded due to a high frequency of missing genotype data (>5%).

616 **Clonality.**

617 Most computational tools for population genetics are based on concepts developed
618 for sexual model organisms. Microbial pathogens are often clonal or partially clonal, and
619 hence require different tools to address their population dynamics and evolutionary history.
620 The R package *poppr 2.0* (Kamvar et al., 2015) specifically addresses issues with analysis of
621 clonal and partially clonal populations. We first used this package to collapse individuals into
622 clonal groups, by defining a genetic distance threshold based on 3 different clustering
623 algorithms using the function `mlg.filter` (Kamvar et al., 2015). This initial step, besides
624 enabling definition of clonal lineage boundaries is a necessary partial correction for a bias
625 that affects metrics of most computational tools that often rely on allele frequencies
626 assuming panmixia. A dissimilarity matrix was produced by *poppr* to compute genetic
627 distances between genomes and a minimum spanning network was drawn based on these
628 calculations, by collapsing individuals based on the previously defined genetic distance
629 threshold. Neighbour-joining trees were generated for each intercontinental lineage with
630 *ape* R package (Paradis and Schliep, 2019) using the dissimilarity matrix produced by *poppr* R
631 package.

632

633 **Global ancestry inference.**

634 In order to identify ancestral populations and to characterize the admixture patterns
635 in our dataset, we used ADMIXTURE 1.3 (Alexander and Lange, 2011). ADMIXTURE is useful

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

636 as an exploratory tool in analyses of genetic structure, but should be interpreted with
637 caution, since such model-based algorithms often provide only a caricature of a complex
638 reality. The dataset was clone-censored (including only one randomly chosen strain from
639 each clonal lineage) and pruned for linkage disequilibrium in PLINK10 (v. 1.07) using
640 parameters --indep-pairwise 50 5 0.2 (it removes each SNP that has an R^2 value greater than
641 0.2 with any other SNP within a 50-SNP sliding window, advanced by 5 SNPs each time).
642 ADMIXTURE was run using the unsupervised mode with cross-validation (--cv). The number
643 of ancestral populations (K) varied between 2 and 10.

644 We complemented our global ancestry analyses with ChromoPainter (Lawson et al.,
645 2012), known to be particularly useful to discern signatures of recent admixture.
646 ChromoPainter estimates the number of "chunks" of ancestry inherited by an individual or a
647 population from a "donor" individual or population, and builds a co-ancestry matrix that
648 summarizes the degree of sharing of ancestry among all pairs of individuals. Unlike
649 ADMIXTURE, ChromoPainter takes into account patterns of linkage disequilibrium allowing
650 one to combine information across successive markers to increase the ability to capture fine-
651 scale population structure. Hence, linkage disequilibrium pruning was not required for this
652 analysis and we used the unpruned clone-censored dataset.

653

654 **Local ancestry inference.**

655 Local ancestry inference (also designated as ancestry deconvolution) is the task of
656 identifying the regional ancestral origin of chromosomal segments in admixed individuals. It
657 requires specifying a set of candidate non-admixed populations as putative ancestors of the
658 admixed individuals. After defining ancestral and admixed individuals using approaches of
659 global ancestry inference we carried out local ancestry analysis using the recently released
660 Ancestry_HMM software (Corbett-Detig and Nielsen, 2017). Ancestry_HMM is based on a
661 novel hidden Markov model that does not require genotypes from reference panels and that
662 is generalized to arbitrary ploidy, and is hence suitable for non-model haploid organisms.

663 In order to estimate the divergence between ancestral populations and their putative
664 progeny as defined by Ancestry_HMM, we generated neighbour-joining trees for each of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

665 13 chromosomes using *ape* R package that computed genetic distances based on the
666 dissimilarity matrix produced by *poppr* R package. In addition, a TCS network was produced
667 from the apicoplast sequences using PopART.

668

669 **Dating emergence, global spread, and introgressions of intercontinental *T. gondii* clonal**
670 **lineages.**

671 We estimated the mutation rate of *T. gondii* based on the *in vivo* mutation rate of the
672 RH strain, through successive passage in outbred mice during 30 years (Supplementary
673 Methods). We estimated the mutation rate of *T. gondii* to range between 3.1×10^{-9} to
674 11.7×10^{-9} mutations per site per year. We used Tang's equation (Tang et al., 2002) to
675 calculate the time to the most recent common ancestor (TMRCA), which we adapted to
676 haploid genomes : $T_2 = \frac{d_{ij}}{\mu * \ell}$, where d_{ij} is the number of nucleotide differences between any
677 two sequences i and j, μ is the mutation rate per site and ℓ is the length of the studied
678 genomic region.

679

680 **Candidate genes for adaptation to the domestic environment.**

681 To identify candidate genes involved in the process of adaptation to the domestic
682 environment a multistage process was used. We first carried out a divergence-based
683 selection scan using the population branch statistic (PBS) (Yi et al., 2010). This method
684 enables to detect genomic regions of unexpectedly high differentiation between different
685 pre-defined groups (estimated with FST measure by Hudson et al. (1992)), a pattern
686 indicative of directional selection (Lewontin and Krakauer, 1973). In parallel, in order to
687 ascertain candidate SNPs, we used the bcftools 1.1 -- *private* function to identify SNPs that
688 differentiate domestic strains from wild strains (alleles common to domestic strains and not
689 found in wild strains). SNPs identified with these two approaches were annotated with
690 SnpEff in order to define a primary list of candidate genes. To determine which candidates
691 genes are most likely to be under positive selection, we focused on functional evidence. We
692 carefully searched the literature for the function of candidate genes and their stage-related

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

693 patterns of expression. Selection can result in patterns of extended linkage disequilibrium
694 (LD) and extended haplotype homozygosity (EHH) around the selected site, especially
695 relative to the alternative allele. EHH is defined as the probability that two randomly chosen
696 chromosomes carrying the same allele at a focal SNP are identical by descent over a given
697 distance surrounding it. Hence, for outlier SNPs validated by all types of evidence —
698 environmental, statistical and functional—, we computed EHH using the using the *REHH 2.0*
699 R package (Gautier et al., 2017) to examine the extent of linkage disequilibrium around
700 them.

701

702 ## Acknowledgements

703 We thank the French Agence Nationale de la Recherche (ANR project IntroTox 17-CE35-0004) and
704 the Nouvelle-Aquitaine region (Directorate of Research, Higher Education and Technology Transfer),
705 for funding this research. We thank Nicolas Plault, Lionel Forestier and Eden Lebrault for technical
706 assistance. We want to acknowledge the Biological Resource Center for *Toxoplasma* for providing
707 strains and the French National Reference Center for toxoplasmosis for data regarding human
708 strains. We thank Aurélien Dumètre and Emmanuelle Gilot-Fromont for their advice on estimating
709 the generation time of *Toxoplasma gondii*. We would also like to thank Gordon Langsley for revision
710 of the English language. Genome sequencing was performed by the genomic platform GENOM'IC of
711 Cochin Institute in Paris. Animal work was conducted in the animal facility managed by the technical
712 platform BISCEm (US042 Inserm - UMS 2015 CNRS) of the University and CHU of Limoges. The
713 computations presented in this article were carried out on the CALI calculator of the University of
714 Limoges (CAlcul en LImousin), funded by the Limousin region, the European Union, the XLIM, IPAM,
715 GEIST institutes, and the University of Limoges.

716

717 ## References

718 Ajzenberg, D., Collinet, F., Mercier, A., Vignoles, P., Dardé, M.-L., 2010. Genotyping of Toxoplasma
719 gondii isolates with 15 microsatellite markers in a single multiplex PCR assay. J. Clin.
720 Microbiol. 48, 4641–4645. https://doi.org/10.1128/JCM.01152-10
721 Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry
722 estimation. BMC bioinformatics 12, 1–6.
723 Bauer, H., Packer, C., Funston, P.F., Henschel, P., Nowell, K., 2016. Panthera leo (errata version
724 published in 2017). The IUCN Red List of threatened species 201.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

725 Bertranpetit, E., Jombart, T., Paradis, E., Pena, H., Dubey, J., Su, C., Mercier, A., Devillard, S.,
726 Ajzenberg, D., 2017. Phylogeography of Toxoplasma gondii points to a South American origin.
727 Infect. Genet. Evol. 48, 150–155. https://doi.org/10.1016/j.meegid.2016.12.020
728 Bonhomme, F., Orth, A., Cucchi, T., Rajabi-Maham, H., Catalan, J., Boursot, P., Auffray, J.-C., Britton-
729 Davidian, J., 2011. Genetic differentiation of the house mouse around the Mediterranean
730 basin: matrilineal footprints of early and late colonization. Proceedings of the Royal Society
731 B: Biological Sciences 278, 1034–1043.
732 Boyle, J.P., Rajasekar, B., Saeij, J.P., Ajioka, J.W., Berriman, M., Paulsen, I., Roos, D.S., Sibley, L.D.,
733 White, M.W., Boothroyd, J.C., 2006. Just one cross appears capable of dramatically altering
734 the population biology of a eukaryotic pathogen like Toxoplasma gondii. Proceedings of the
735 National Academy of Sciences 103, 10514–10519.
736 Calarco, L., Barratt, J., Ellis, J., 2018. Genome wide identification of mutational hotspots in the
737 apicomplexan parasite Neospora caninum and the implications for virulence. Genome
738 biology and evolution 10, 2417–2431.
739 Carme, B., Demar, M., Ajzenberg, D., Dardé, M., 2009. Severe acquired toxoplasmosis caused by wild
740 cycle of Toxoplasma gondii, French Guiana [WWW Document]. Emerging infectious diseases.
741 https://doi.org/10.3201/eid1504.081306
742 Chaichan, P., Mercier, A., Galal, L., Mahittikorn, A., Ariey, F., Morand, S., Boumédiène, F., Udonsom,
743 R., Hamidovic, A., Murat, J.B., Sukthana, Y., Dardé, M.L., 2017. Geographical distribution of
744 Toxoplasma gondii genotypes in Asia: A link with neighboring continents. Infect. Genet. Evol.
745 https://doi.org/10.1016/j.meegid.2017.06.002
746 Corbett-Detig, R., Nielsen, R., 2017. A hidden Markov model approach for simultaneously estimating
747 local ancestry and admixture time using next generation sequence data in samples of
748 arbitrary ploidy. PLoS genetics 13, e1006529.
749 Di Genova, B.M., Wilson, S.K., Dubey, J.P., Knoll, L.J., 2019. Intestinal delta-6-desaturase activity
750 determines host range for Toxoplasma sexual reproduction. PLoS biology 17, e3000364.
751 Driscoll, C.A., Menotti-Raymond, M., Roca, A.L., Hupe, K., Johnson, W.E., Geffen, E., Harley, E.H.,
752 Delibes, M., Pontier, D., Kitchener, A.C., Yamaguchi, N., O'brien, S.J., Macdonald, D.W., 2007.
753 The Near Eastern origin of cat domestication. Science 317, 519–523.
754 https://doi.org/10.1126/science.1139518
755 Elbez-Rubinstein, A., Ajzenberg, D., Dardé, M.-L., Cohen, R., Dumètre, A., Yera, H., Gondon, E.,
756 Janaud, J.-C., Thulliez, P., 2009. Congenital toxoplasmosis and reinfection during pregnancy:
757 case report, strain characterization, experimental model of reinfection, and review. J. Infect.
758 Dis. 199, 280–285. https://doi.org/10.1086/595793
759 Farhat, D.C., Swale, C., Dard, C., Cannella, D., Ortet, P., Barakat, M., Sindikubwabo, F., Belmudes, L.,
760 De Bock, P.-J., Couté, Y., 2020. A MORC-driven transcriptional switch controls Toxoplasma
761 developmental trajectories and sexual commitment. Nature microbiology 5, 570–583.
762 Fitak, R.R., Rinkevich, S.E., Culver, M., 2018. Genome-wide analysis of SNPs is consistent with no
763 domestic dog ancestry in the endangered Mexican wolf (Canis lupus baileyi). Journal of
764 Heredity 109, 372–383.
765 Froese, D., Stiller, M., Heintzman, P.D., Reyes, A.V., Zazula, G.D., Soares, A.E., Meyer, M., Hall, E.,
766 Jensen, B.J., Arnold, L.J., 2017. Fossil and genomic evidence constrains the timing of bison
767 arrival in North America. Proceedings of the National Academy of Sciences 114, 3457–3462.
768 Galal, L., Ajzenberg, D., Hamidović, A., Durieux, M.-F., Dardé, M.-L., Mercier, A., 2018. Toxoplasma
769 and Africa: One Parasite, Two Opposite Population Structures. Trends in Parasitology 34,
770 140–154. https://doi.org/10.1016/j.pt.2017.10.010
771 Galal, L., Hamidović, A., Dardé, M., Mercier, A., 2019. Diversity of Toxoplasma gondii strains at the
772 global level and its determinants [WWW Document]. Food and waterborne parasitology.
773 https://doi.org/10.1016/j.fawpar.2019.e00052
774 Galal, L., Sarr, A., Cuny, T., Brouat, C., Coulibaly, F., Sembène, M., Diagne, M., Diallo, M., Sow, A.,
775 Hamidović, A., Plault, N., Dardé, M.-L., Ajzenberg, D., Mercier, A., 2019. The introduction of

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

776 new hosts with human trade shapes the extant distribution of Toxoplasma gondii lineages.
777 PLoS Negl Trop Dis 13, e0007435. https://doi.org/10.1371/journal.pntd.0007435
778 Galal L, Stragier C, Boumédiène F, Hamidović A, Maugrion O, Dardé Ml, Mercier A, 2020. Combining
779 spatial analysis and host population genetics to gain insights into the mode of transmission of
780 a pathogen: The example of Toxoplasma gondii in mice [WWW Document]. Infection,
781 genetics and evolution : journal of molecular epidemiology and evolutionary genetics in
782 infectious diseases. https://doi.org/10.1016/j.meegid.2019.104142
783 Gautier, M., Klassmann, A., Vitalis, R., 2017. rehh 2.0: a reimplementation of the R package rehh to
784 detect positive selection from haplotype structure. Molecular ecology resources 17, 78–90.
785 Gel, B., Serra, E., 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes
786 displaying arbitrary data. Bioinformatics 33, 3088–3090.
787 Goodrich, J., Lynam, A., Miquelle, D., Wibisono, H., Kawanishi, K., Pattanavibool, A., Htun, S., Tempa,
788 T., Karki, J., Jhala, Y., 2015. Panthera tigris. The IUCN Red List of Threatened Species 2015: e.
789 T15955A50659951.
790 Gotteland, C., Gilot-Fromont, E., Aubert, D., Poulle, M.-L., Dupuis, E., Dardé, M.-L., Forin-Wiart, M.-A.,
791 Rabilloud, M., Riche, B., Villena, I., 2014. Spatial distribution of Toxoplasma gondii oocysts in
792 soil in a rural area: Influence of cats and land use. Veterinary Parasitology 205, 629–637.
793 Guthrie, R.D., 2004. Radiocarbon evidence of mid-Holocene mammoths stranded on an Alaskan
794 Bering Sea island. Nature 429, 746–749.
795 Hamidović, A., Etougbétché, J.R., Tonouhewa, A.B.N., Galal, L., Dobigny, G., Houémènou, G., Da
796 Zoclanclounon, H., Amagbégnon, R., Laleye, A., Fievet, N., 2021. A hotspot of Toxoplasma
797 gondii Africa 1 lineage in Benin: How new genotypes from West Africa contribute to
798 understand the parasite genetic diversity worldwide. PLoS neglected tropical diseases 15,
799 e0008980.
800 Henn, B.M., Botigué, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua,
801 P.A., Moreno-Estrada, A., Bertranpetit, J., 2012. Genomic ancestry of North Africans supports
802 back-to-Africa migrations. PLoS Genet 8, e1002397.
803 Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA
804 sequence data. Genetics 132, 583–589.
805 Innes, E.A., Hamilton, C., Garcia, J.L., Chryssafidis, A., Smith, D., 2019. A one health approach to
806 vaccines against Toxoplasma gondii. Food and Waterborne Parasitology 15, e00053.
807 Jewell, M.L., Frenkel, J.K., Johnson, K.M., Reed, V., Ruiz, A., 1972. Development of Toxoplasma
808 oocysts in neotropical felidae. The American journal of tropical medicine and hygiene 21,
809 512–517.
810 Jiang, T., Shwab, E., Martin, R., Gerhold, R., Rosenthal, B., Dubey, J., Su, C., 2018. A partition of
811 Toxoplasma gondii genotypes across spatial gradients and among host species, and
812 decreased parasite diversity towards areas of human settlement in North America [WWW
813 Document]. International journal for parasitology.
814 https://doi.org/10.1016/j.ijpara.2018.01.008
815 Jones, E.P., Eager, H.M., Gabriel, S.I., Jóhannesdóttir, F., Searle, J.B., 2013. Genetic tracking of mice
816 and other bioproxies to infer human history. Trends in Genetics 29, 298–308.
817 Jones, E.P., Skirnisson, K., McGovern, T.H., Gilbert, M.T.P., Willerslev, E., Searle, J.B., 2012. Fellow
818 travellers: a concordance of colonization patterns between mice and men in the North
819 Atlantic region. BMC Evolutionary Biology 12, 1–8.
820 Kamvar, Z.N., Brooks, J.C., Grünwald, N.J., 2015. Novel R tools for analysis of genome-wide
821 population genetic data with emphasis on clonality. Frontiers in genetics 6, 208.
822 Khan, A., Ajzenberg, D., Mercier, A., Demar, M., Simon, S., Dardé, M.L., Wang, Q., Verma, S.K.,
823 Rosenthal, B.M., Dubey, J.P., Sibley, L.D., 2014a. Geographic separation of domestic and wild
824 strains of Toxoplasma gondii in French Guiana correlates with a monomorphic version of
825 chromosome1a. PLoS Negl Trop Dis 8, e3182. https://doi.org/10.1371/journal.pntd.0003182

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

826 Khan, A., Dubey, J.P., Su, C., Ajioka, J.W., Rosenthal, B.M., Sibley, L.D., 2011. Genetic analyses of
827 atypical Toxoplasma gondii strains reveal a fourth clonal lineage in North America. Int. J.
828 Parasitol. 41, 645–655. https://doi.org/10.1016/j.ijpara.2011.01.005
829 Khan, A., Fux, B., Su, C., Dubey, J.P., Darde, M.L., Ajioka, J.W., Rosenthal, B.M., Sibley, L.D., 2007.
830 Recent transcontinental sweep of Toxoplasma gondii driven by a single monomorphic
831 chromosome. Proc. Natl. Acad. Sci. U.S.A. 104, 14872–14877.
832 https://doi.org/10.1073/pnas.0702356104
833 Khan, A., Shaik, J.S., Behnke, M., Wang, Q., Dubey, J.P., Lorenzi, H.A., Ajioka, J.W., Rosenthal, B.M.,
834 Sibley, L.D., 2014b. NextGen sequencing reveals short double crossovers contribute
835 disproportionately to genetic diversity in Toxoplasma gondii. BMC genomics 15, 1–15.
836 Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O.A., Jang, J., Sung, S., Lee, S., Salim, B., Jung, J., 2020.
837 The mosaic genome of indigenous African cattle as a unique genetic resource for African
838 pastoralism. Nature Genetics 52, 1099–1110.
839 Konstantinovic, N., Guegan, H., Stäjner, T., Belaz, S., Robert-Gangneux, F., 2019. Treatment of
840 toxoplasmosis: Current options and future perspectives. Food and Waterborne Parasitology
841 15, e00036.
842 Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of population structure using dense
843 haplotype data. PLoS Genet 8, e1002453.
844 Lehmann, T., Marcet, P.L., Graham, D.H., Dahl, E.R., Dubey, J.P., 2006. Globalization and the
845 population structure of Toxoplasma gondii. Proc. Natl. Acad. Sci. U.S.A. 103, 11423–11428.
846 https://doi.org/10.1073/pnas.0601438103
847 Lewontin, R.C., Krakauer, J., 1973. Distribution of gene frequency as a test of the theory of the
848 selective neutrality of polymorphisms. Genetics 74, 175–195.
849 Lipinski, M.J., Froenicke, L., Baysac, K.C., Billings, N.C., Leutenegger, C.M., Levy, A.M., Longeri, M.,
850 Niini, T., Ozpinar, H., Slater, M.R., 2008. The ascent of cat breeds: genetic evaluations of
851 breeds and worldwide random-bred populations. Genomics 91, 12–21.
852 Lorenzi, H., Khan, A., Behnke, M.S., Namasivayam, S., Swapna, L.S., Hadjithomas, M., Karamycheva,
853 S., Pinney, D., Brunk, B.P., Ajioka, J.W., Ajzenberg, D., Boothroyd, J.C., Boyle, J.P., Dardé, M.L.,
854 Diaz-Miranda, M.A., Dubey, J.P., Fritz, H.M., Gennari, S.M., Gregory, B.D., Kim, K., Saeij, J.P.J.,
855 Su, C., White, M.W., Zhu, X.-Q., Howe, D.K., Rosenthal, B.M., Grigg, M.E., Parkinson, J., Liu, L.,
856 Kissinger, J.C., Roos, D.S., Sibley, L.D., 2016. Local admixture of amplified and diversified
857 secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. Nat
858 Commun 7, 10147. https://doi.org/10.1038/ncomms10147
859 Lowe, J.J., Walker, M., 2014. Reconstructing quaternary environments. Routledge.
860 Macholán, M., Baird, S.J., Munclinger, P., Piálek, J., 2012. Evolution of the house mouse. Cambridge
861 University Press.
862 Mercier, A., Ajzenberg, D., Devillard, S., Demar, M.P., de Thoisy, B., Bonnabau, H., Collinet, F.,
863 Boukhari, R., Blanchet, D., Simon, S., Carme, B., Dardé, M.-L., 2011. Human impact on genetic
864 diversity of Toxoplasma gondii: example of the anthropized environment from French
865 Guiana. Infect. Genet. Evol. 11, 1378–1387. https://doi.org/10.1016/j.meegid.2011.05.003
866 Mercier, A., Devillard, S., Ngoubangoye, B., Bonnabau, H., Bañuls, A.-L., Durand, P., Salle, B.,
867 Ajzenberg, D., Dardé, M.-L., 2010. Additional haplogroups of Toxoplasma gondii out of Africa:
868 population structure and mouse-virulence of strains from Gabon. PLoS Negl Trop Dis 4, e876.
869 https://doi.org/10.1371/journal.pntd.0000876
870 Miller, N.L., Frenkel, J.K., Dubey, J.P., 1972. Oral infections with Toxoplasma cysts and oocysts in
871 felines, other mammals, and in birds. The Journal of parasitology 928–937.
872 Milne, G., Webster, J.P., Walker, M., 2020. Toxoplasma gondii: an underestimated threat? Trends in
873 parasitology.
874 Minot, S., Melo, M.B., Li, F., Lu, D., Niedelman, W., Levine, S.S., Saeij, J.P., 2012. Admixture and
875 recombination among Toxoplasma gondii lineages explain global genome diversity.
876 Proceedings of the National Academy of Sciences 109, 13458–13463.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

877 Montoya, J.G., Liesenfeld, O., 2004. Toxoplasmosis. Lancet 363, 1965–1976.
878 https://doi.org/10.1016/S0140-6736(04)16412-X
879 Müller, U.B., Howard, J.C., 2016. The impact of Toxoplasma gondii on the mammalian genome.
880 Current opinion in microbiology 32, 19–25.
881 Ottoni, C., Van Neer, W., De Cupere, B., Daligault, J., Guimaraes, S., Peters, J., Spassov, N.,
882 Prendergast, M.E., Boivin, N., Morales-Muñiz, A., 2017. The palaeogenetics of cat dispersal in
883 the ancient world. Nature Ecology & Evolution 1, 1–7.
884 Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
885 analyses in R. Bioinformatics 35, 526–528.
886 Peters, J., 1998. Römische Tierhaltung und Tierzucht: eine Synthese aus archäozoologischer
887 Untersuchung und schriftlich-bildlicher Überlieferung. Leidorf.
888 Phillips, L., Chambers, N., Backensto, S., 2018. Migration: On the move in Alaska. Alaska Park Science
889 17, 1–3.
890 Pomares, C., Devillard, S., Holmes, T.H., Olariu, T.R., Press, C.J., Ramirez, R., Talucod, J., Estran, R., Su,
891 C., Dubey, J.P., 2018. Genetic characterization of Toxoplasma gondii DNA samples isolated
892 from humans living in North America: an unexpected high prevalence of atypical genotypes.
893 The Journal of infectious diseases 218, 1783–1791.
894 Puckett, E.E., Park, J., Combs, M., Blum, M.J., Bryant, J.E., Caccone, A., Costa, F., Deinum, E.E., Esther,
895 A., Himsworth, C.G., 2016. Global population divergence and admixture of the brown rat
896 (Rattus norvegicus). Proceedings of the Royal Society B: Biological Sciences 283, 20161762.
897 Ramakrishnan, C., Maier, S., Walker, R.A., Rehrauer, H., Joekel, D.E., Winiger, R.R., Basso, W.U., Grigg,
898 M.E., Hehl, A.B., Deplazes, P., 2019. An experimental genetically attenuated live vaccine to
899 prevent transmission of Toxoplasma gondii by cats. Scientific reports 9, 1–14.
900 Ribeiro, A., Golicz, A., Hackett, C.A., Milne, I., Stephen, G., Marshall, D., Flavell, A.J., Bayer, M., 2015.
901 An investigation of causes of false positive single nucleotide polymorphisms using simulated
902 reads from a small eukaryote genome. BMC bioinformatics 16, 1–16.
903 Schumacher, A., Elbadawi, L., DeSalvo, T., Straily, A., Ajzenberg, D., Letzer, D., Moldenhauer, E.,
904 Handly, T., Hill, D., Dardé, M., Pomares, C., Passebosc-Faure, K., Bisgard, K., Gomez, C., Press,
905 C., Smiley, S., Montoya, J., Kazmierczak, J., 2020. Toxoplasmosis Outbreak Associated With
906 Toxoplasma gondii-Contaminated Venison-High Attack Rate, Unusual Clinical Presentation,
907 and Atypical Genotype [WWW Document]. Clinical infectious diseases : an official publication
908 of the Infectious Diseases Society of America. https://doi.org/10.1093/cid/ciaa285
909 Shwab, E.K., Saraf, P., Zhu, X.-Q., Zhou, D.-H., McFerrin, B.M., Ajzenberg, D., Schares, G., Hammond-
910 Aryee, K., van Helden, P., Higgins, S.A., 2018. Human impact on the diversity and virulence of
911 the ubiquitous zoonotic parasite Toxoplasma gondii. Proceedings of the National Academy of
912 Sciences 115, E6956–E6963.
913 Shwab, E.K., Zhu, X.-Q., Majumdar, D., Pena, H.F.J., Gennari, S.M., Dubey, J.P., Su, C., 2014.
914 Geographical patterns of Toxoplasma gondii genetic diversity revealed by multilocus PCR-
915 RFLP genotyping. Parasitology 141, 453–461. https://doi.org/10.1017/S0031182013001844
916 Su, C., Evans, D., Cole, R.H., Kissinger, J.C., Ajioka, J.W., Sibley, L.D., 2003. Recent expansion of
917 Toxoplasma through enhanced oral transmission. Science 299, 414–416.
918 https://doi.org/10.1126/science.1078035
919 Su, C., Khan, A., Zhou, P., Majumdar, D., Ajzenberg, D., Dardé, M.-L., Zhu, X.-Q., Ajioka, J.W.,
920 Rosenthal, B.M., Dubey, J.P., Sibley, L.D., 2012. Globally diverse Toxoplasma gondii isolates
921 comprise six major clades originating from a small number of distinct ancestral lineages.
922 Proc. Natl. Acad. Sci. U.S.A. 109, 5844–5849. https://doi.org/10.1073/pnas.1203190109
923 Subrahmanyam, S., Alam, M., 2007. Indo-Persian Travels In the Age of Discoveries. Cambridge:
924 Cambridge University Press.
925 Tang, H., Siegmund, D.O., Shen, P., Oefner, P.J., Feldman, M.W., 2002. Frequentist estimation of
926 coalescence times from nucleotide sequence data using a tree-based partition. Genetics 161,
927 447–459.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.457674; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

928 Torgerson, P.R., Mastroiacovo, P., 2013. The global burden of congenital toxoplasmosis: a systematic
929 review. Bull. World Health Organ. 91, 501–508. https://doi.org/10.2471/BLT.12.111732
930 VanWormer, E., Conrad, P.A., Miller, M.A., Melli, A.C., Carpenter, T.E., Mazet, J.A., 2013. Toxoplasma
931 gondii, source to sea: higher contribution of domestic felids to terrestrial parasite loading
932 despite lower infection prevalence. EcoHealth 10, 277–289.
933 Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N.,
934 Korneliussen, T.S., 2010. Sequencing of 50 human exomes reveals adaptation to high
935 altitude. science 329, 75–78.
936