Thèse de doctorat

Université de Limoges

École Doctorale Gay Lussac - Sciences pour l'Environnement (ED 523)

UNITE DE RECHERCHE UMR INSERM 1094 Neuroépidémiologie tropicale (NET)

Thèse pour obtenir le grade de

Docteur de l'Université de Limoges

Discipline/Spécialité : Parasitologie

Présentée et soutenue par

Emilie Bertranpetit

Le 19 décembre 2016

Sur l'origine de *Toxoplasma gondii* : approches phylogénétique et spatialement-explicite pour la détermination de l'origine géographique d'un parasite ubiquiste

Thèse dirigée par Dr. Daniel Ajzenberg et Dr. Sébastien Devillard

JURY :

Président du jury

M. Pierre Flori, Professeur, laboratoire de Parasitologie-Mycologie, CHU de Saint-Etienne-Hôpital Nord

Rapporteurs

M. Pierre Flori, Professeur, laboratoire de biologie-pathologie, CHU de Saint-Etienne-Hôpital Nord

M. Ignacio G. Bravo, Directeur de recherche, laboratoire MIVEGEC, Université de Montpellier

Examinateurs

M. Sébastien Devillard, Maître de conférences, laboratoire de Biométrie et de Biologie Evolutive, Université de Lyon

M. Daniel Ajzenberg, Maître de conférences, laboratoire de Parasitologie-Mycologie, Université de Limoges

Remerciements

Je remercie mes directeurs de thèse, les Dr. Daniel AJZENBERG et Dr. Sébastien DEVILLARD pour avoir encadré ce travail durant ces trois années ainsi que pour leurs conseils et leur disponibilité.

Je remercie également le Pr. Pierre FLORI ainsi que le Dr. Ignacio G. BRAVO pour avoir accepté de faire partie du jury de cette thèse.

Je remercie toutes les personnes qui ont permis la réalisation de ce travail en nous fournissant des souches ainsi que les Dr. Thibaut JOMBART et Dr. Emmanuel PARADIS pour les analyses effectuées durant cette thèse.

Je voudrais aussi remercier le Pr. Pierre-Marie PREUX ainsi que le Pr. Marie-Laure DARDÉ pour m'avoir accueillie dans leur laboratoire et m'avoir permis de réaliser ce travail.

Merci au Dr. Aurélien MERCIER pour toutes ses discussions et ses idées concernant ce travail.

Je tiens aussi à remercier Martine GATET pour m'avoir appris les différentes techniques de laboratoire dont j'avais besoin durant ma thèse ainsi que Roselyne MOUZET pour son aide lors des séances de TP. Merci à toutes les deux pour votre disponibilité et votre bonne humeur.

Merci au Dr. Bertrand COURTIOUX pour m'avoir permis d'effectuer quelques heures d'enseignement et au Dr. Jean-Benjamin MURAT pour ses conseils.

Merci à tous les autres membres de l'équipe pour leur aide et leur bonne humeur notamment Patcharee, Mayoura, Clotilde, Julien, Paula, Lokman, Nicolas et tous les autres.

Droits d’auteurs

Cette création est mise à disposition selon le Contrat :
« **Attribution-Pas d'Utilisation Commerciale-Pas de modification 3.0 France** »
disponible en ligne : *http://creativecommons.org/licenses/by-nc-nd/3.0/fr/*



Sommaire

Introduction....5
Chapitre I. Classification, biologie et histoire évolutive de *Toxoplasma gondii*....6
I.1. Place de *Toxoplasma gondii* parmi les Apicomplexes....6
I.2. Cycle de vie de *T. gondii* et diversité génétique....10
I.3. Etat de l'art sur l'histoire évolutive de *T. gondii* et de ses hôtes définitifs....22
I.4. Objectifs de la thèse....33
Chapitre II. Etude expérimentale....37
II.1. Le jeu de données....37
II.2. Diversité génétique de *T. gondii*....48
II.3. Phylogénie....58
II.4. Analyse spatiale....72
II.5. Datations....76
II.6. Discussion générale....81
Perspectives et conclusion....84
Références bibliographiques....85
Annexes....95

Introduction

Le parasitisme est une interaction durable dans laquelle un organisme, l'hôte, représente le milieu de vie d'un autre organisme, le parasite. Le parasite exploite l'hôte pour sa survie et sa reproduction ce qui entraîne des effets plus ou moins délétères sur celles de l'hôte. Il s'agit d'un des modes de vie présentant le plus de succès et qui est apparu dans de nombreux groupes phylogénétiques (Poulin and Morand, 2000).
Malgré les avancées médicales et l'amélioration de l'hygiène, les maladies infectieuses représentent encore 30% des principales causes de morbidité et de mortalité humaine ainsi que 25% des décès dans le monde, les zones les plus touchées étant les régions tropicales. Environ 58 à 61 % des pathogènes humains sont considérés comme zoonotiques et parmi eux 73 à 75 % seraient émergents (Smith and Guégan, 2010). Dans ce contexte où la prévalence et la fréquence des maladies infectieuses zoonotiques augmentent, il apparait important de mieux comprendre la distribution géographique, aussi bien en termes de variabilité génétique que d'incidence, et l'histoire évolutive des pathogènes zoonotiques (Han et al., 2016).
Découvert simultanément en 1908 chez un rongeur d'Afrique du Nord, *Ctenodactylus gundi,* par Nicole et Manceaux en Tunisie et chez un lapin au Brésil par Splendore (Ferguson, 2009), *Toxoplasma gondii* est un parasite protozoaire intracellulaire obligatoire, responsable de la toxoplasmose. L'Homme peut s'infecter par l'ingestion d'oocystes excrétés par les félidés qui sont les hôtes définitifs ainsi que par l'ingestion de kystes tissulaires lors de la consommation de viande mal cuite provenant des animaux homéothermes qui sont les hôtes intermédiaires. Bien qu'asymptomatique chez les personnes immunocompétentes, cette maladie peut avoir de graves conséquences chez les personnes immunodéprimées et chez les fœtus lorsque la mère s'est infectée en cours de grossesse. De plus, il existe une hétérogénéité géographique des formes cliniques de la toxoplasmose, avec la présence de certaines formes sévères particulièrement en Amérique du Sud.
La structure de population de ce parasite est complexe et peut s'expliquer en partie par certaines de ces caractéristiques telles qu'un large spectre d'hôtes, la présence d'un cycle sexué chez ses hôtes définitifs ainsi que sa capacité à se transmettre d'hôte intermédiaire en hôte intermédiaire via l'ingestion de kystes tissulaires par carnivorisme. Des variations géographiques de la diversité génétique existent avec des régions présentant une diversité élevée, et d'autres régions où prédominent seulement quelques génotypes clonaux majoritaires. L'histoire évolutive de ce parasite fait débat notamment son origine géographique. La diversité génétique plus élevée en Amérique du Sud tendrait à indiquer qu'il s'agit de l'origine géographique de *T. gondii*. Cependant certaines datations moléculaires et la mise en relation avec l'histoire évolutive de ses hôtes définitifs suggèrent plutôt qu'il serait entré en Amérique du Sud via les migrations des félidés. La connaissance de cette origine est un élément indispensable pour comprendre l'évolution d'un tel parasite ubiquiste et donc le lien qu'il peut y avoir entre les formes cliniques de la toxoplasmose et la diversité génétique de *T. gondii*.
Dans ce travail, nous proposons de déterminer l'origine géographique de *T. gondii* à l'aide d'une analyse spatiale ainsi que de déterminer la date du plus récent ancêtre commun des souches actuelles de ce parasite (TMRCA). Des analyses phylogénétiques ont également été réalisées pour déterminer les relations de parenté entre les différentes souches à l'aide des mêmes marqueurs que l'analyse spatiale.

Chapitre I. Classification, biologie et histoire évolutive de *Toxoplasma gondii*

I.1. Place de *Toxoplasma gondii* parmi les Apicomplexes

I.1.1. Les Apicomplexes

Les Apicomplexes sont des parasites intracellulaires obligatoires, dont certains sont importants en santé humaine et animale (Morrison, 2009). Parmi les plus connus, on peut citer les genres *Plasmodium*, *Eimeria* ou encore *Theileria* ainsi que *Toxoplasma gondii,* responsable respectivement du paludisme, de coccidioses animales, de la theiléríose et de la toxoplasmose. Le paludisme est une maladie transmise par les moustiques du genre *Anopheles* dans les régions pauvres, tropicales et subtropicales du monde, et est une des principales causes de mortalité depuis le début de l'Humanité, en particulier chez les enfants (Tanabe et al., 2010). L'Afrique sub-saharienne paye le plus lourd tribut à cette maladie où les jeunes enfants de moins de 5 ans et les femmes enceintes sont les plus vulnérables (White et al., 2014). La toxoplasmose est une maladie bénigne le plus souvent mais peut avoir de graves conséquences chez les personnes immunodéprimées et chez les fœtus lors d'infections congénitales. L'incidence globale annuelle de la toxoplasmose congénitale a été estimée à environ 200 000 cas (Torgerson and Mastroiacovo, 2013). Cette maladie est un problème de santé publique en Amérique du Sud à cause de la forte prévalence de formes oculaires dans la population générale de certaines régions du Brésil, de Colombie et d'Argentine (Glasner et al., 1992; Rudzinski et al., 2016; de-la-Torre et al., 2009) et de la possibilité de formes sévères potentiellement fatales à des personnes immunocompétentes en Guyane française (Carme et al., 2002; Demar et al., 2007, 2012). Les espèces du genre *Eimeria* et *Theileria* nuisent quant à elles à la santé animale et entraînent des pertes économiques importantes. Le genre *Eimeria* provoque des coccidioses chez les principaux animaux d'élevage et sont en terme d'impact économique et de bien-être animal les parasites les plus importants chez les poulets domestiques (Clark et al., 2016). Les espèces du genre *Theileria* infectent un grand nombre d'animaux à la fois sauvages et domestiques. *T. parva*, *T. annulata* et *T. lestoquardi* sont par exemple des espèces pathogènes pour les ruminants domestiques (Nene et al., 2016).

L'invasion des cellules hôtes par les Apicomplexes se fait de façon active par mouvement de glissement, grâce à un moteur moléculaire appelé le glidéosome, basé sur un système actine-myosine. *T. gondii* et *Plasmodium* sont les parasites de ce groupe les plus étudiés pour comprendre les mécanismes d'invasion. Le cycle lytique du toxoplasme se déroule en plusieurs étapes. La première consiste à l'attachement du parasite à la cellule hôte par des antigènes de surface, suivie par une réorientation permettant au pôle apical de se placer face à la surface de la cellule hôte. Ce pôle apical comprend le complexe apical qui est une caractéristique des Apicomplexes. Il est constitué du conoïde, et de deux organites de sécrétion, les micronèmes et les rhoptries, dont les protéines sécrétées interviennent au moment de l'invasion (Figure 1). Le parasite entre alors dans la cellule et une vacuole parasitophore se forme. Elle joue notamment un rôle de protection, en permettant au parasite de se multiplier à l'abri du système immunitaire de l'hôte, avant de se disséminer (Frénal and Soldati-Favre, 2013). Un autre organite, l'apicoplaste est présent chez la plupart des Apicomplexes, et est caractéristique de ce groupe (Figure 1). Il s'agit d'un plaste sans activité photosynthétique et présentant un génome d'environ 30-40 kb (Arisue and Hashimoto, 2015), hérité de façon maternelle (Ferguson et al., 2005). Cet apicoplaste aurait évolué à partir de

BY-NC-ND

l'endosymbiose secondaire d'une algue rouge. Les Apicomplexes qui ne possèdent pas ce plaste, tel que le genre *Cryptosporidium*, l'auraient perdu de façon secondaire (Arisue and Hashimoto, 2015).

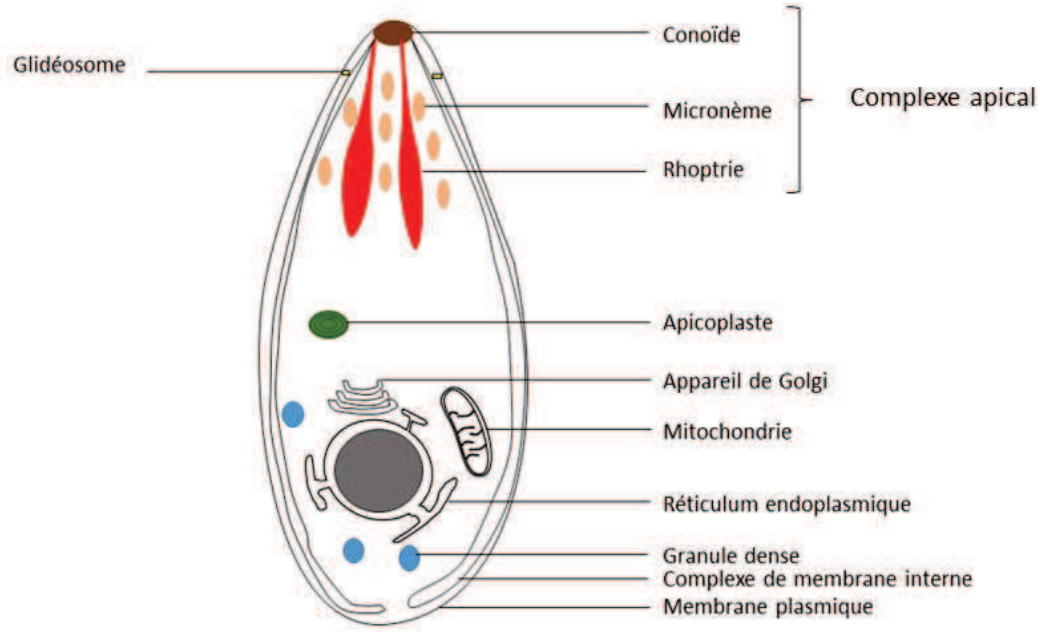


Figure 1 : Schéma de la structure d'un tachyzoïte de *Toxoplasma gondii.*

Source : Adapté de Frénal and Soldati-Favre (2013)

Malgré leur importance médicale et vétérinaire, ce groupe est mal connu. Avec environ 6000 espèces identifiées sur les 1,2 à 10 x 10^6 espèces potentielles (Adl et al., 2007), il fait partie des groupes d'organismes du vivant les moins bien connus (Morrison, 2009). Cela est en grande partie dû au fait que l'échantillonnage en parasitologie est souvent opportuniste, et se concentre donc sur les parasites étudiés dans les cas de maladies humaines et animales (Morrison, 2009). Or, bien que les Apicomplexes soient tous parasites, seuls quelques-uns sont pathogènes pour l'Homme et les animaux domestiques, ce qui explique que de rares espèces responsables de maladies humaines et animales (toxoplasmose, paludisme, cryptosporidiose, coccidiose…) soient parmi les plus étudiées dans le monde tandis que la majorité des autres espèces sont parmi les plus méconnues (Lorenzi et al., 2016; Morrison, 2009). Il s'agit d'un biais important en reconstruction phylogénétique puisque l'absence de nombreuses espèces ne permet pas de détecter toutes les relations de parenté, alors qu'il serait important d'avoir une phylogénie la plus complète possible de ce groupe afin de mieux les étudier (Morrison, 2009).

I.1.2. Classification des Apicomplexes

Traditionnellement les Apicomplexes sont divisés en 4 groupes : les coccidies, les grégarines, les hémosporidies et les piroplasmes. Cette classification est utilitaire et repose majoritairement sur des critères phénotypiques tels que par exemple le cycle de vie ou des caractères morphologiques et ultra-structuraux (Morrison, 2009; Tenter et al., 2002). Bien que ces caractères soient utiles pour la description des espèces, il est difficile d'y trouver de

BY-NC-ND

véritables homologies et donc des informations phylogénétiquement utiles (Tenter et al., 2002). Les caractères moléculaires, utilisés par la suite, sont considérés comme plus informatifs (Tenter et al., 2002). Cependant, la phylogénie des Apicomplexes reste incertaine. Outre le biais d'échantillonnage déjà évoqué, d'autres paramètres expliquent cette absence de phylogénie stable : les caractères moléculaires utilisés, le choix des analyses phylogénétiques, et les discordances entre les classifications (Morrison, 2009).

Les phylogénies des Apicomplexes qui s'appuient sur des caractères génétiques reposent essentiellement sur le gène de la petite sous-unité de l'ARN ribosomique (ARN ribosomique 18S). Une phylogénie basée sur un seul gène risque de refléter la phylogénie du gène et non pas de l'espèce. Cependant dans le cas des Apicomplexes, les autres gènes disponibles proviennent souvent d'études sur d'autres thématiques telles que les vaccins et sont impliqués dans la reconnaissance de l'hôte et dans l'interaction avec le système immunitaire. Ces gènes ont donc tendance à être soumis à une forte pression de sélection ou à être spécifiques d'un groupe rendant la reconstruction phylogénétique non-robuste ou impossible (Morrison, 2009). Cependant les groupes monophylétiques déterminés en utilisant le gène ARNr 18S coïncident assez bien avec les grands groupes suggérés par les caractères morphologiques et le cycle de vie (Tenter et al., 2002).

La figure ci-dessous (Figure 2) présente une phylogénie réalisée avec un autre gène (DEAD box helicase protein) (Lorenzi et al., 2016). Elle illustre les 4 grands groupes traditionnellement trouvés. *Toxoplasma gondii* appartient au groupe des Sarcocystidae ou coccidies formant des kystes tissulaires, caractéristique qu'il partage avec *Hammondia hammondii*, *Neospora caninum* et *Sarcocystis neurona*.

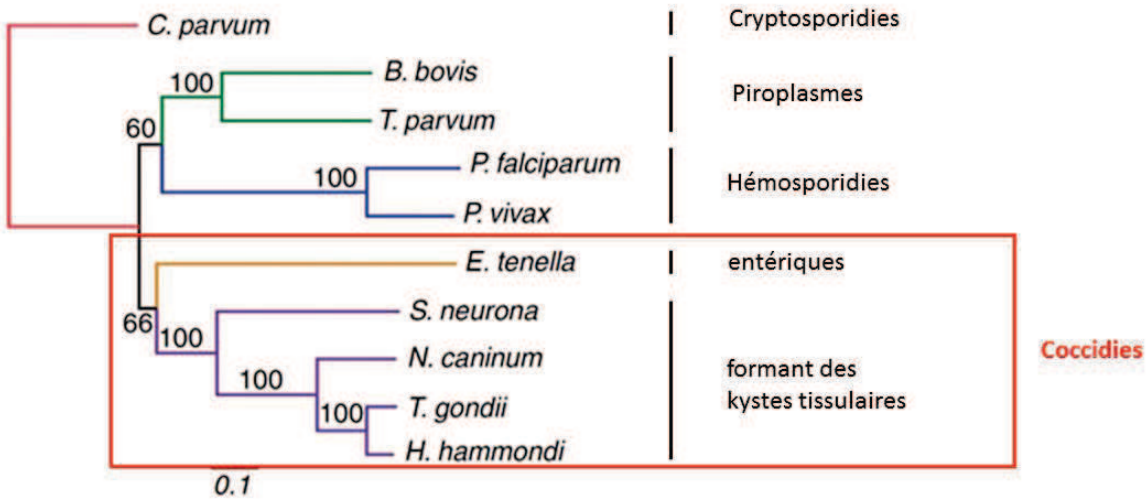


Figure 2 : Arbre phylogénétique (Neighbour joining) illustrant les 4 grands groupes d'Apicomplexes.

Source : Adapté de Lorenzi et al. (2016)

BY-NC-ND

Toxoplasma gondii possède un génome d'environ 65 Mb constitué de 14 chromosomes et de 8322 gènes codant pour des protéines. La comparaison de génome avec ceux d'autres Apicomplexes a permis de mettre en évidence que le degré de conservation de l'ordre des gènes sur les chromosomes ou synténie était très élevé entre *T. gondii* et *H. hammondi*, et ne diminuait que faiblement lorsqu'ils étaient comparés à *Neospora caninum*. Cependant, la synténie est faible voir absente lorsque c'est *Sarcocystis neurona* et *Eimeria tenella* qui sont comparés aux trois premiers parasites (Lorenzi et al., 2016). Ces résultats confortent le fait que *T. gondii*, *H. hammondi* et *N. caninum* sont plus proches entre eux qu'avec les autres parasites du même groupe et d'ailleurs ils font partie d'une même famille, les Toxoplasmatinae (Morrison, 2009; Morrison et al., 2004; Tenter et al., 2002).

I.1.3. Relations de parenté dans la sous-famille des Toxoplasmatinae

Outre les genres *Neospora*, *Hammondia* et *Toxoplasma*, certaines espèces du genre *Isospora* feraient également partie des Toxoplasmatinae car ce genre est polyphylétique (Morrison, 2009). Les genres *Toxoplasma*, *Neospora* et *Hammondia* sont très proches et forment un clade (Morrison et al., 2004; Tenter et al., 2002). Il a été montré que le genre *Hammondia* n'était pas monophylétique. En effet, *Hammondia heydorni* serait plus proche de *N. caninum* tandis que *H. hammondi* serait plus proche de *T. gondii*.
Les dates de divergence de ces différentes espèces sont assez controversées et difficiles à estimer en raison du manque de connaissances sur la phylogénie de ce groupe. Une phylogénie de quelques Apicomplexes (*E. tenella*, *Sarcocystis tenella*, *S. neurona*, *Isospora belli*, *H. hammondi*, *N. caninum* et *T. gondii*) a été réalisée avec les régions de la petite sous-unité (SSU) et de l'ITS1 de l'ADN ribosomique. Cette phylogénie ne résout pas les relations de parenté entre *H. hammondi*, *N. caninum* et *T. gondii* et montre une trifurcation pour ces trois parasites. Le nœud correspondant à cette trifurcation a été estimé à 12 Ma en utilisant un taux moyen de substitution de la région SSU calculé à partir de plusieurs taxons (Su et al., 2003). Une autre phylogénie incluant un nombre plus important d'espèces du groupe des coccidies et s'appuyant sur des séquences de l'ARN ribosomique 18S montre que *T. gondii* est plus proche d'*H. hammondi* que de *N. caninum* (Morrison et al., 2004). Avec un taux moyen de substitution de l'ARNr 18S de 1% pour 100 millions d'années, le TMRCA (Time to the Most Recent Common Ancestor ou date du plus récent ancêtre commun) de *N. caninum* et *T. gondii* a été estimé à 11,3 Ma. Cependant *N. caninum*, *H. hammondi* et *T. gondii* semblent très proches sur cette phylogénie avec moins de différences nucléotidiques entre ces trois espèces de genres différents qu'entre certaines espèces du même genre, comme par exemple deux espèces de *Besnoitia* (Morrison et al., 2004). En résumé, l'apparition de l'espèce *T. gondii* est estimée à environ 11 Ma d'après la datation précédente mais la séparation des trois espèces *N. caninum*, *H. hammondi* et *T. gondii* ne semble pas très claire car les séquences utilisées dans cette étude présentent peu de polymorphismes pour distinguer les trois parasites (Morrison, 2005). Une autre étude comparant les deux génomes complets de *N. caninum* et *T. gondii*, en appliquant la théorie de l'horloge moléculaire, a estimé la divergence entre ces deux espèces à environ 28 Ma avec un intervalle de confiance compris entre 21,7 et 42,7 Ma (Reid et al., 2012).
Ces différentes études montrent que le choix des séquences est crucial pour reconstruire la phylogénie des Apicomplexes. Les analyses sur génomes complets permettent d'avoir plus d'informations que celles reposant sur un gène. La datation faite à partir des génomes complets de *T. gondii* et *N. caninum* donne une estimation plus ancienne de leur divergence. Il est donc possible que les séquences sur l'ADN ribosomal sous-estiment les dates de

BY-NC-ND

divergence. Cependant, la date de divergence entre *T. gondii* et *H. hammondi* est plus récente que celle entre *T. gondii* et *N. caninum* et se situerait donc aux alentours de 11 Ma.

I.2. Cycle de vie de *T. gondii* et diversité génétique

I.2.1. Cycle de vie

I.2.1.1. Cycle de vie de *T. gondii* et comparaison avec *H. hammondi*

T. gondii présente un large spectre d'hôtes contrairement à beaucoup d'autres Apicomplexes (Grigg and Sundar, 2009), avec les félidés comme hôtes définitifs et les homéothermes comme hôtes intermédiaires (Figure 3).
La reproduction sexuée a lieu dans l'épithélium intestinal du Félidé. Il y a différentiation entre gamète mâle et gamète femelle, fécondation puis excrétion d'oocystes dans l'environnement via les fécès. Un chat infecté peut excréter des millions d'oocystes (Dubey, 2001). Ces oocystes ne deviennent infectant qu'après sporulation. Ils contiennent alors 8 sporozoïtes et peuvent souiller l'eau, le sol ou les végétaux ce qui permet leur ingestion par un autre hôte. Les sporozoïtes contenus dans les oocystes sont alors libérés et se convertissent en tachyzoïtes, qui se répliquent dans la vacuole parasitophore. Ils sortent ensuite activement de la cellule hôte en la détruisant et envahissent de nouvelles cellules. Cette succession d'étapes invasion, réplication, sortie constitue le cycle lytique du toxoplasme. Les tachyzoïtes sont les formes impliquées dans la dissémination rapide du parasite lors de la phase aigüe de l'infection. Sous le contrôle de la réponse immunitaire de l'hôte, les tachyzoïtes se différentient ensuite en bradyzoïtes, formes dormantes, contenues dans des kystes tissulaires qui marquent la phase chronique de l'infection. Ces bradyzoïtes ont la possibilité de se transformer en tachyzoïtes, notamment lors d'une faiblesse du système immunitaire de l'hôte. C'est ce processus qui explique la réactivation de l'infection acquise de façon antérieure par les personnes immunodéprimées. Les bradyzoïtes peuvent conduire à une infection lorsqu'ils sont ingérés par un autre hôte, définitif ou intermédiaire (Blader et al., 2015). La transmission d'hôte intermédiaire en hôte intermédiaire par carnivorisme est une particularité de *T. gondii* qui permet une propagation clonale du parasite. Lorsque l'infection est acquise au cours d'une grossesse, les tachyzoïtes peuvent aussi traverser le placenta et infecter le fœtus en cours de développement. Trois voies de contamination existent donc : celle par les oocystes de l'environnement, celle par les kystes tissulaires et la transmission congénitale.

Bien qu'ils aient divergé il y a environ 11 Ma, *T. gondii* et *H. hammondi* sont morphologiquement proches et jusqu'en 1975, ils ont été confondus. Cependant des différences majeures existent dans leur cycle de vie (Figures 3 et 4). En ce qui concerne le spectre d'hôtes, celui d'*H. hammondi* semble plus restreint bien que des incertitudes demeurent, principalement en raison du faible nombre d'études sur ce parasite. Par exemple, les hôtes définitifs sont les chats, et probablement l'ensemble des félidés comme c'est le cas pour *T. gondii* mais il est difficile de trouver une étude rapportant la présence d'*H. hammondi* chez des félidés sauvages. Toutefois, une étude portant sur 45 échantillons de fécès provenant de félidés sauvages en Amérique Centrale a rapporté la présence d'oocystes ressemblant à ceux de *T. gondii* mais n'a pas pu exclure la possibilité qu'il s'agisse en fait d'oocystes d'*H. hammondi* par exemple (Patton et al., 1986). Les hôtes intermédiaires naturels trouvés sont les souris, les rats, les chevreuils et les chèvres (Dubey et al., 2013). D'autres hôtes, testés de façon expérimentale, ont été déterminés. Il s'agit des chiens, des cochons, des singes et des lapins mais contrairement à *T. gondii* les oiseaux ne font pas partie des

BY-NC-ND

hôtes intermédiaires d'*H. hammondi* (Dubey and Sreekumar, 2003). Une autre différence entre les deux parasites est qu'*H. hammondi* a obligatoirement un cycle à deux hôtes. Les oocystes et les tachyzoïtes n'infectant pas les chats, ces derniers ne peuvent donc s'infecter qu'en ingérant des proies présentant des kystes tissulaires contenant des bradyzoïtes d*'H. hammondi*. En ce qui concerne les hôtes intermédiaires, seuls les oocystes peuvent les infecter. De plus, *H. hammondi* ne se transmet pas de façon congénitale. Ces caractéristiques rendent obligatoire le passage par l'hôte définitif et par l'hôte intermédiaire et ne donnent pas la possibilité à ce parasite de se propager d'hôte intermédiaire en hôte intermédiaire par carnivorisme, comme c'est le cas pour *T. gondii*. Chez le chat, il n'y a pas de développement extra intestinal, et seuls des schizontes et des gamontes ont été détectés. Chez la souris contrairement à *T. gondii*, les kystes d'*H. hammondi* se forment préférentiellement au niveau des muscles et des poumons et non pas du cerveau. (Dubey and Sreekumar, 2003). Le faible nombre d'études sur *H. hammondi* s'explique principalement par le fait qu'il s'agit d'un organisme qui ne provoque pas de manifestations cliniques. Il est de plus, difficile à différentier de *T. gondii*. En ce qui concerne les examens coprologiques, les oocystes d'*H. hammondi* et de *T. gondii* sont similaires. Il existe aussi des réactions croisées en sérologie mais aussi en biologie moléculaire selon les amorces utilisées (Dubey et al., 2013). L'inoculation aux souris reste donc l'un des meilleurs moyens de les différencier et ceci en raison des différences de leur cycle de vie. Les bradyzoïtes de *T. gondii* étant infectant pour les hôtes intermédiaires, ce parasite peut se maintenir indéfiniment chez la souris alors qu'*H. hammondi* ne peut se maintenir plus d'un passage (Dubey and Sreekumar, 2003).

BY-NC-ND

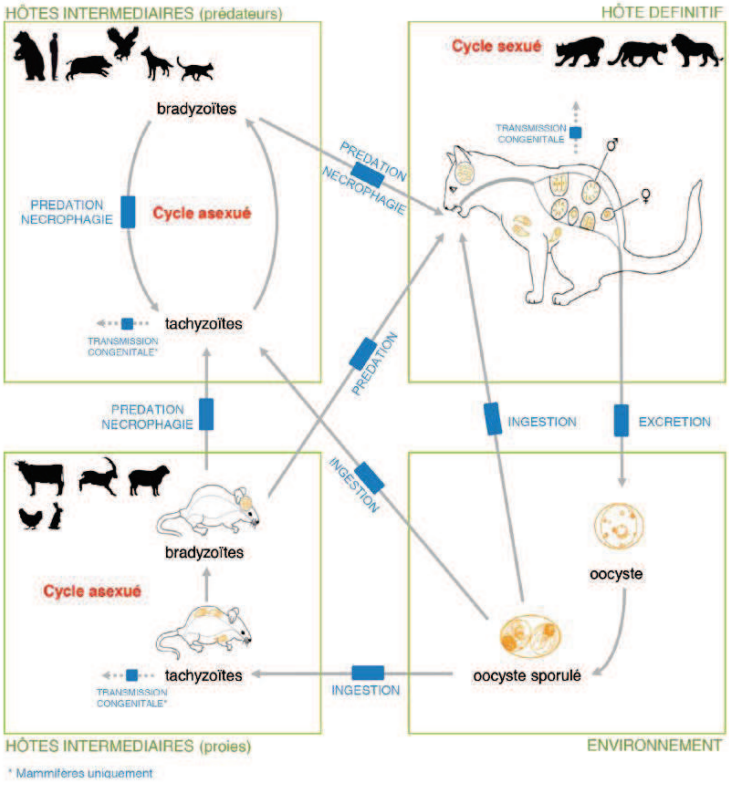


Figure 3 : Cycle de vie de *Toxoplasma gondii*

Source : Mercier (2010). Avec l'autorisation d'Aurélien Mercier.

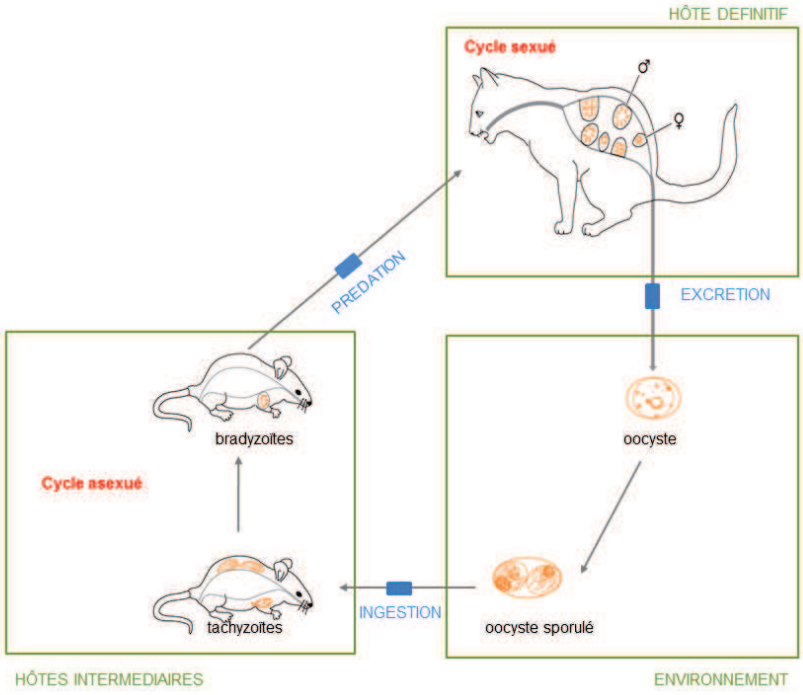


Figure 4 : Cycle de vie d'*Hammondia hammondi*

Source : Adapté de Sibley and Ajioka (2008) et de Mercier (2010)

BY-NC-ND

I.2.2. Diversité génétique de *T. gondii*

Plusieurs arguments biologiques, principalement des différences entre isolats de *T. gondii* concernant la virulence chez la souris, la capacité de production d'oocystes et la vitesse de réplication en culture cellulaire ont fait soupçonner l'existence d'une certaine diversité génétique dans la population du parasite (Dardé, 1997). Dans les années 1990, des techniques permettant l'étude de la structure de population de *T. gondii* se sont développées (Ferguson, 2009).

I.2.2.1. Marqueurs utilisés pour génotyper *T. gondii*

L'objectif des mises au point des différentes méthodes de génotypage est d'avoir une résolution suffisamment élevée pour distinguer les différentes souches.
Les premières études ont utilisé les isoenzymes et la méthode de PCR-RFLP qui ont montré des résultats concordants (Dardé et al., 1992; Howe et al., 1997).
Actuellement le génotypage est principalement réalisé à l'aide de trois méthodes multilocus, la prise en compte d'un seul locus ne permettant pas de déterminer avec précision le génotype d'un isolat :

- Les microsatellites correspondent à des répétitions d'un court motif nucléotidique (1 à 6 bp) et sont omniprésents dans les génomes eucaryotes. Ces répétitions entraînent des erreurs de la polymérase lors de la réplication, conduisant à un nombre de répétitions différents. Le polymorphisme de longueur induit est alors utilisé dans cette méthode pour différencier les souches (Ajzenberg et al., 2010).
- Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) : méthode basée sur des endonucléases qui reconnaissent des single nucleotide polymorphisms (SNPs) et coupent le produit PCR à cet endroit. Après cette digestion, les fragments sont révélés par électrophorèse sur gel d'agarose. Il s'agit donc encore d'un polymorphisme de longueur (Su et al., 2009).
- Le séquençage de fragments d'ADN : cette méthode permet d'avoir accès à l'ensemble de la séquence et donc à différents types de polymorphismes : SNPs, insertions et délétions.

En général, le taux de mutation des microsatellites chez les eucaryotes est estimé entre 10^{-2} et 10^{-5} par locus par réplication (Su et al., 2009). Il est donc plus important que celui des SNPs identifiés avec les deux dernières méthodes qui est d'environ 10^{-9} à 10^{-10} par position nucléotidique et par réplication. Ces différences dans les taux de mutations peuvent améliorer la résolution de la méthode basée sur les microsatellites, la rendant particulièrement utile pour différentier des isolats génétiquement très proches. Le séquençage est long et couteux mais est indispensable en cas d'étude phylogénétique. Malgré son faible pouvoir discriminatoire, la technique PCR-RFLP est la plus répandue dans les laboratoires en raison de son faible coût car elle ne nécessite pas d'appareil de séquençage.
La comparaison des résultats entre les différentes techniques est un problème car actuellement 3 systèmes de désignation de génotypes existent en lien avec l'utilisation de marqueurs précis pour les 3 méthodes ci-dessus.

Une méthode par microsatellite utilisant une seule PCR multiplex avec 15 marqueurs localisés sur 11 chromosomes différents (*TUB2, W35, TgM-A, B18, B17, M33, IV.1, XI.1, M48, M102, N60, N82, AA, N61, N83*) a été mise au point. Elle permet deux niveaux de discrimination des souches. Le premier, avec les 8 marqueurs *TUB2, W35, TgM-A, B18, B17, M33, IV.1* et *XI.1*

BY-NC-ND

permet le typage des souches, en différenciant les 3 groupes clonaux majeurs I, II et III et les génotypes atypiques. Parmi ces génotypes atypiques, certains sont caractéristiques de régions géographiques et ont une dénomination qui rappelle cette origine, comme par exemple Africa 1–3, Caribbean 1–3 ou Chinese 1. Le deuxième niveau avec les marqueurs *M48, M102, N60, N82, AA, N61, N83* permet une discrimination plus fine en différentiant des souches appartenant à la même lignée (Ajzenberg et al., 2010). La méthode par PCR-RFLP repose sur une PCR multiplex avec 10 marqueurs (*SAG1, SAG2, SAG3, BTUB, GRA6, c22-8, c29-2, L358*, PK1 et *Apico*). Pour résoudre le problème de la désignation de nombreux génotypes, ceux identifiés par PCR-RFLP sont désignés comme des génotypes « ToxoDB PCR-RFLP #» suivis d'un nombre spécifique. La liste des 231 génotypes décrits à l'heure actuelle par PCR-RFLP est disponible sur une base de données publique (http://toxodb.org/toxo/showQuestion.do?questionFullName=RflpIsolateQuestions.ByGenotypeNumber, accédé le 17/10/2016). La désignation par haplogroupes est quant à elle basée sur la méthode du séquençage multilocus de 8 introns de 5 loci : *UPRT* (introns 1 et 7), *MIC2* (introns 1 et 3), *BTUB* (introns 1 et 2), *EF1* (intron1) et *HP* (intron 2). Cependant, d'autres marqueurs peuvent être associés à ces introns comme des marqueurs PCR-RFLP (Su et al., 2012). De plus, l'évolution des méthodes de séquençage a également conduit à utiliser les SNPs provenant de génomes complets afin d'identifier les haplogroupes (Lorenzi et al., 2016). A l'heure actuelle, 16 haplogroupes ont été décrits.

I.2.2.2. Principaux résultats

Ces méthodes de génotypage permettent d'identifier de nouveaux génotypes et de décrire leur circulation dans les différentes régions. Cependant, il est difficile à partir de ces résultats de comprendre les liens entre les différentes souches. Il faut souligner aussi l'important biais d'échantillonnage de toutes les études qui privilégient l'isolement de souches provenant d'animaux domestiques (poulets, chats, chiens, moutons, etc.) dans des zones géographiques accessibles. La collection de souches de *T. gondii* est difficile car le parasite est tissulaire, ce qui nécessite l'inoculation du parasite encore vivant aux souris à partir de tissus frais et donc la présence d'une animalerie et d'une logistique de terrain particulièrement efficace. Ces limitations expliquent la rareté des souches provenant d'animaux sauvages ou d'endroits difficiles d'accès sans infrastructure adaptée. Quand on regarde la carte de provenance des souches de *T. gondii*, il est clair que des régions immenses comme l'Afrique sub-saharienne ou l'Asie sont largement sous-échantillonnées, ce qui est un biais majeur dans notre connaissance de la diversité de ce parasite (Figure 5).

Les premières études réalisées tendaient à suggérer que *T. gondii* présentait une structure de population clonale avec 3 principaux types clonaux (I, II et III). L'utilisation de méthodes comprenant un nombre croissant de marqueurs ainsi que l'augmentation de l'échantillonnage dans des zones reculées ont permis de mettre à jour une structure plus complexe. Les génotypes définis par les méthodes de PCR-RFLP et microsatellite ont permis de déterminer quels génotypes étaient présents et dominaient dans différentes régions géographiques. Jusqu'en 2012, le typage par PCR-RFLP a permis de classer 1457 isolats en 189 génotypes et a révélé que la diversité génétique de *T. gondii* se structurait géographiquement. Un contraste existe entre l'Amérique du Sud et Centrale et les autres régions. En effet, dans la plupart des régions, un ou plusieurs génotypes dominent. L'Amérique du Nord présenterait 40 génotypes mais les génotypes ToxoDB PCR-RFLP #1, 2, 3, 4 et 5 prédomineraient. Le génotype #1 correspond au type clonal II, le #2 au type III, le #3 au type II variant, les #4 et #5

BY-NC-ND

au type 12 (E Keats Shwab, 2013). Le type 12 est un génotype clonal d'abord détecté chez les loutres de Californie puis chez les animaux sauvages en Amérique du Nord (Khan et al., 2011a; Miller et al., 2004; Sundar et al., 2008). En Europe 9 génotypes ont été détectés mais les génotypes #1, 2 et 3 prédomineraient. L'Asie présenterait 10 génotypes avec une dominance des génotypes #9 et #10. Le génotype #10 correspond au type clonal I tandis que le génotype #9 correspond à *Chinese 1*. En Afrique où 13 génotypes ont été détectés, ce sont les génotypes #2 et 3 qui seraient majoritaires. En Amérique du Sud et Centrale, 156 génotypes étaient présents et aucun ne dominerait de façon claire. Contrairement aux autres régions, les types clonaux I, II et III s'y trouveraient en faible fréquence. Cependant, des contrastes existent entre les différentes zones, comme le montre la prédominance des types II et III au Chili, séparé de la plupart des autres pays par la cordillère des Andes (Figure 5). Les diversités intra-populationnelles ont montré une plus forte diversité en Amérique du Sud et Centrale, suivies par l'Amérique du Nord. Les divergences entre populations ont été mesurées à l'aide de l'indice Fst. Le plus petit Fst a été calculé entre l'Amérique du Nord et l'Europe pour lesquelles aucune différence significative n'a été trouvée. Des différences significatives ont été trouvées entre toutes les autres populations, avec le plus fort Fst détecté entre l'Amérique du Nord d'une part et l'Amérique du Sud et Centrale d'autre part (E Keats Shwab, 2013).

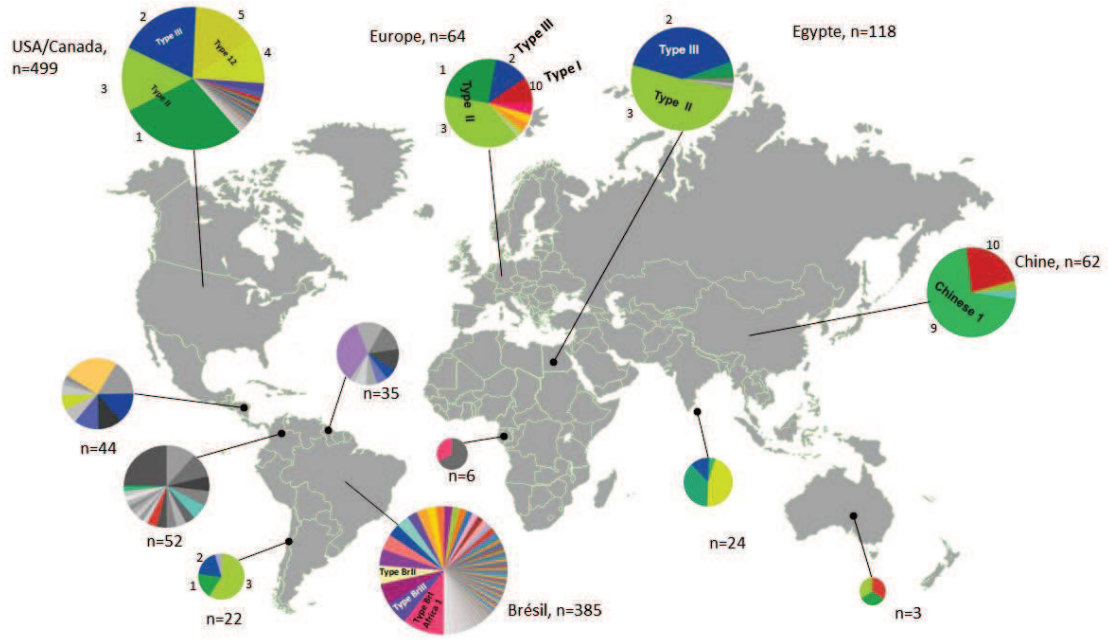


Figure 5 : Distribution géographique des génotypes de *Toxoplasma gondii*. Les chiffres autour des diagrammes en secteurs représentent les génotypes ToxoDB PCR-RFLP. Dans certains cas, pour les principaux génotypes, le nom alternatif a été donné en gras. La taille des diagrammes est corrélée au nombre d'échantillons (n) pris en compte dans la zone. Les couleurs différentes représentent des génotypes différents.

Source : Adapté de E Keats Shwab (2013)

Les analyses conduisant à la formation de réseaux, ont permis de définir des haplogroupes et des clades, qui donnent des indications basiques sur les relations entre les génotypes. L'identification des haplogroupes évolue avec l'ajout de nouvelles souches. Onze

BY-NC-ND

haplogroupes ont d'abord été définis à partir de 46 souches sur la base des 8 introns de 5 locus. Parmi eux, les groupes 1 à 3 étaient prédominants en Europe et en Amérique du Nord, les haplogroupes 4, 5 et 8-10 rassemblaient essentiellement des souches d'Amérique du Sud. Le groupe 6 était largement répandu (Europe, Amérique du Sud et Afrique) (Khan et al., 2007). La prise en compte d'isolats d'Amérique du Nord provenant d'hôtes domestiques mais aussi sauvages et ayant été génotypés en type X ou A, et en atypiques, a conduit à la détermination de l'haplogroupe 12 sur la base des 8 introns. Cette analyse a été confirmée par l'utilisation de gènes codant pour un antigène de surface (*SAG1*) et pour des protéines de granules denses (*GRA6* et *GRA7*) (Khan et al., 2011a). L'addition de souches chinoises et africaines a permis ensuite de définir 2 nouveaux haplogroupes, le 13 et le 14. L'haplogroupe 13 est constitué de souches chinoises, dont le génotype avait déjà été défini par PCR-RFLP. L'analyse des introns montre que ces souches forment un haplogroupe indépendant. L'haplogroupe 14 est constitué de souches génotypées comme *Africa 3* par l'analyse microsatellite (Khan et al., 2011b). L'haplogroupe 15 fut ensuite défini. Il est formé de souches brésiliennes, et est proche des haplogroupes 5 et 10 dont les souches ont été isolées en Guyane Française (Su et al., 2012). Un seizième haplogroupe a été proposé à partir d'une analyse sur les SNPs du génome (Figure 6). Certaines souches qui se groupaient avec les haplogroupes 4 et 8 se démarquent nettement dans cette analyse (Lorenzi et al., 2016). Cette désignation en haplogroupes n'est donc pas fixe et change en fonction des méthodes et des souches utilisées.

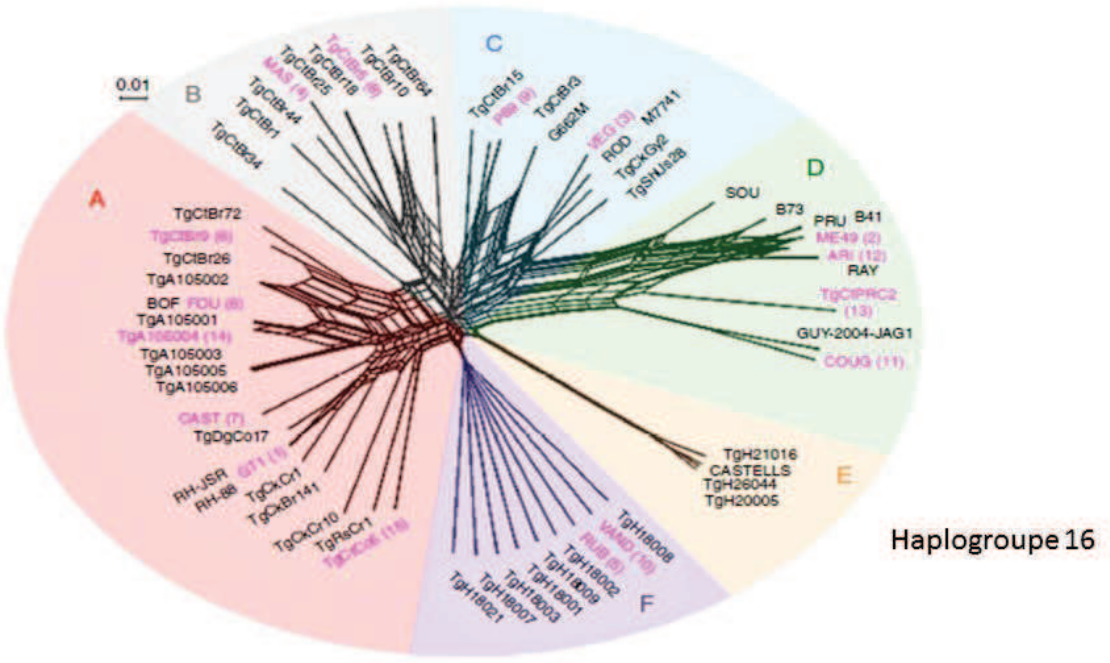


Figure 6 : Structure de population de *Toxoplasma gondii*. Analyse en Neighbor-net, réalisée à partir de SNPs provenant de génomes complets. Les couleurs de fond avec les lettres correspondent aux principaux clades. Les haplogroupes sont indiqués par les chiffres en face des souches représentatives Barre d'échelle : nombre de SNPs par site.

Source : Adapté de Lorenzi et al.(2016)

BY-NC-ND

I.2.2.3. Les moteurs de la diversité génétique de *T. gondii*

La complexité du cycle de *T. gondii* ainsi que les différentes études sur sa diversité génétique ont conduit à un débat sur la part de la propagation clonale et des recombinaisons génétiques dans la structure de population de ce pathogène (Wendte et al., 2011). Ce débat s'étend plus largement aux pathogènes protozoaires (Ramírez and Llewellyn, 2014; Tibayrenc and Ayala, 2014). L'existence d'un cycle sexué bien défini chez les félidés permet la création de diversité génétique. La fécondation a lieu entre un microgamète et un macrogamète dans l'hôte définitif et produit un zygote diploïde. La méiose intervient rapidement après cette fécondation, au stade oocyste, rétablissant l'haploïdie de *T. gondii*. Au cours de cette méiose, les recombinaisons génétiques sont permises par deux mécanismes : l'assortiment indépendant, correspondant à la séparation au hasard des chromosomes homologues et le crossing-over, correspondant à l'échange de matériel génétique par des chromatides non-sœurs de deux chromosomes homologues.
Cependant, dans le cas de *T. gondii*, plusieurs facteurs modulent l'importance de la reproduction sexuée et de la production de diversité génétique, en favorisant une propagation clonale.
Même si le parasite passe par son hôte définitif et qu'il y a reproduction sexuée, l'infection des félidés par une seule souche ne peut produire que des gamètes mâles et femelles génétiquement identiques, conduisant à une fécondation de type « self-mating ». Il a été montré que le « self-mating » était une adaptation clé, pouvant conduire à l'expansion clonale de tels génotypes lors d'une épidémie (Wendte et al., 2010). Bien qu'il s'agisse d'une reproduction sexuée, il n' y a pas de nouvelles combinaisons alléliques dans la progénie puisque les deux gamètes parentaux fournissent des allèles identiques (Sibley and Ajioka, 2008). Il faut donc que deux souches différentes soient ingérées par un Félidé (une proie infectée par deux souches ou ingestion de deux proies différentes à peu de temps d'intervalle) pour que la recombinaison puisse créer de nouvelles combinaisons d'allèles, ce qui semble rare chez *T. gondii*. Il semble donc que ce soit la propagation clonale qui soit favorisée chez ce parasite, cette dernière étant largement amplifiée par le fait que les kystes puissent infecter les hôtes intermédiaires par carnivorisme et qu'il y ait possibilité de transmission congénitale (Grigg and Sundar, 2009). Les premières études sur la diversité de *T. gondii* suggéraient d'ailleurs qu'il présentait une structure de population clonale avec les trois principaux types I, II et III. Cependant, le sur-échantillonnage de quelques génotypes ne permet pas d'exclure une structure sous-jacente comprenant divers génotypes présentant des taux d'échanges génétiques élevés, comme c'est le cas pour les structures de population épidémiques (Wendte et al., 2011). Les 3 principaux types clonaux seraient d'ailleurs des recombinants, soulignant l'importance du rôle de la reproduction sexuée dans la population de *T. gondii* (Boyle et al., 2006). La diversification de l'échantillonnage a également conduit à reconsidérer l'hypothèse d'une population clonale, en mettant en évidence la présence d'une diversité et de recombinaisons génétiques plus importantes que ce qui était pensé, dans certaines régions en particulier en Amérique du Sud (Pena et al., 2008; Vitaliano et al., 2014) mais aussi, dans une moindre mesure en Afrique (Bontell et al., 2009; Mercier et al., 2010) et en Asie (Dubey et al., 2007b, 2007a). Les observations faites à partir d'animaux domestiques et d'humains ne sont pas les mêmes que celles faites à partir d'animaux sauvages (Ajzenberg et al., 2004; Vitaliano et al., 2014), amenant à parler de l'existence d'un cycle domestique et d'un cycle sylvatique. Une étude réalisée en Guyane Française où un environnement anthropisé le long des côtes coexiste avec l'environnement sauvage de la forêt amazonienne, a permis de mettre en évidence deux populations génétiquement distinctes. Par rapport à la population

BY-NC-ND

anthropisée, la population sauvage présenterait une diversité plus élevée, pas de déséquilibre de liaison et n'était pas structurée. La richesse en hôtes intermédiaires et la présence de plusieurs espèces de félidés dans la forêt amazonienne influenceraient les flux de gènes et la structure de population de *T. gondii* dans cet environnement. Les félidés seraient plus fréquemment infectés simultanément par des souches différentes qui recombineraient (Mercier et al., 2011). La diversité en mammifères plus importante en forêt amazonienne pourrait expliquer qu'il soit plus avantageux pour le parasite de passer par la reproduction sexuée qui pourrait lui procurer d'avantage de possibilités d'adaptations (Carme et al., 2009). Au contraire, dans l'environnement anthropisé, les transports par l'Homme et la fragmentation des habitats seraient deux forces opposées qui expliqueraient la présence de peu de génotypes mais présentant des fréquences élevées à une échelle locale (Mercier et al., 2011). Dans le cycle domestique, des introgressions fréquentes et des expansions clonales seraient les forces majoritaires tandis que dans le cycle sylvatique, *T. gondii* aurait une population panmictique et la recombinaison serait une force efficace (Grigg and Sundar, 2009). Un modèle de prépondérance de l'évolution clonale a été proposé notamment pour *T. gondii*. Il fait référence à une structure de population, ne nie pas l'existence de recombinaisons mais les considère comme rares et inclut le « self-mating » comme des cas particuliers de clonalité (Tibayrenc and Ayala, 2014). Ce modèle a été remis en question, le manque d'évidence d'échanges génétiques pour la plupart des parasites étant attribué à la difficulté de détecter les indices de la reproduction sexuée plutôt qu'à son absence, notamment en raison d'un échantillonnage de souches inapproprié négligeant les réservoirs non humains et d'un manque de résolution des marqueurs moléculaires utilisés (Ramírez and Llewellyn, 2014).

I.2.3. Séroprévalence et formes cliniques de la toxoplasmose avec un focus sur l'Amérique du Sud

Toxoplasma gondii est donc un parasite ubiquiste qui présente une structure de population complexe avec une diversité génétique qui varie en fonction des zones géographiques. Les formes cliniques de la toxoplasmose sont également diverses. On distingue en général 3 types de toxoplasmoses en fonction du statut immunitaire de l'hôte : i) la toxoplasmose chez l'immunocompétent en général asymptomatique, ii) celle chez l'immunodéprimé qui se présente la plupart du temps sous une forme cérébrale grave due à une réactivation, iii) la toxoplasmose congénitale due à une première infection lors de la grossesse et qui peut avoir plusieurs conséquences pour le fœtus, allant de la mort fœtale à des séquelles neuro-oculaires chez les enfants. Plusieurs études de prévalence de l'infection par *T. gondii*, en particulier chez les femmes enceintes, ont été réalisées afin d'aider à la mise en place des politiques de santé dans les différents pays (Pappas et al., 2009). Ainsi certaines régions d'Amérique du Sud, d'Afrique et d'Asie du Sud-Est présenteraient des taux de séropositivité supérieurs à 60% (Pappas et al., 2009).

L'Amérique du Sud qui présente une diversité génétique des souches plus élevée, montre également des particularités en ce qui concerne l'épidémiologie de la toxoplasmose. Une revue de Dubey et al. (2012) focalisée sur le Brésil relève des séroprévalences en population générale allant de 21,5% à 84% en fonction des régions, avec seulement deux prévalences en dessous des 30%. La comparaison de deux groupes de militaires, l'un au Brésil et l'autre aux Etats-Unis montre que la prévalence au Brésil (56%) est 4 fois plus élevée qu'aux Etats-Unis (13%) (Dubey et al., 2012). Des variations géographiques existent donc entre les pays mais aussi dans un même pays. Plusieurs facteurs peuvent expliquer ces différences comme

BY-NC-ND

les caractéristiques de l'environnement, les pratiques alimentaires, la classe socioéconomique ou le niveau d'étude. En Colombie, les séroprévalences entre deux groupes de militaires, l'un opérant dans la jungle amazonienne et l'autre en milieu urbain à Bogota diffèrent de façon importante (80% et 45% respectivement) (Gómez-Marín et al., 2012). Dans cette étude la consommation d'eau non traitée et la consommation de viande sauvage ont été cités comme des facteurs de risque d'infection toxoplasmique. Une autre étude portant sur des femmes enceintes dans l'état de Paraná au Brésil a mis en évidence une prévalence de 51,7% et les facteurs de risque suivants : habiter en zone rurale, avoir un faible revenu, un faible niveau d'éducation, avoir un âge plus avancé et être en contact régulier avec le sol (Lopes-Mori et al., 2013). Les facteurs de risques peuvent être différents en fonction de la population étudiée. Le manque d'information peut également être mis en cause. En effet, sur 400 femmes enceintes ou venant d'accoucher dans l'état de Rio de Janeiro, 72,2% ne connaissaient pas la toxoplasmose. Parmi ces femmes, la séroprévalence était de 62,3% contre 48,7% parmi celles qui avaient déjà entendu parler de la maladie (Millar et al., 2014).

La séroprévalence est donc élevée en Amérique du Sud mais les études de prévalence chez les femmes enceintes dans cette région montrent qu'il reste néanmoins une part importante de femmes susceptible d'être infectées en cours de grossesse. La charge de morbidité de la toxoplasmose congénitale serait la plus élevée en Amérique du Sud tropicale, en particulier au Brésil, d'après une étude globale effectuée par l'OMS (Torgerson and Mastroiacovo, 2013). Les chiffres d'incidence de la toxoplasmose congénitale en Amérique du Sud varient en fonction des différentes études selon la région et la population étudiée. A Porte Alegre, dans l'état du Rio Grande do Sul au sud du Brésil, l'incidence des toxoplasmoses congénitales était d'environ 0,6/1000 naissances entre les années 2004 et 2014 avec une proportion de formes oculaires voisine de 60% à la naissance (Bischoff et al., 2016). A Belo Horizonte dans l'état du Minas Gerais, entre 2006 et 2007, les auteurs ont rapporté une incidence plus forte (1,3 pour 1000 naissances) et un taux d'atteinte oculaire de 80% après une médiane de suivi de seulement deux mois (Vasconcelos-Santos et al., 2009). En Colombie, une étude a confirmé 15 cas de toxoplasmose congénitale à partir de 15 333 échantillons de sang de cordon, soit environ 1 cas pour 1000 naissances entre 2009 et 2010. Parmi ces 15 cas, 3 (20%) avaient une forme oculaire à la naissance (Gómez-Marin et al., 2011). A titre de comparaison, en France entre 2007 et 2014 l'incidence des toxoplasmoses congénitales était environ de 0,3/1000 naissances (CNR Toxoplasmose, 2015). Les études sur la cohorte lyonnaise d'enfants atteints de toxoplasmose congénitale suggèrent que la prévalence des lésions oculaires en France serait d'environ 30% après une médiane de suivi de 10,5 ans (Wallon et al., 2014). En plus de leurs fréquences plus élevées, les formes cliniques au Brésil semblent aussi être plus graves. Les lésions oculaires chez des enfants avec toxoplasmose congénitale du Brésil semblent plus sévères que chez des enfants d'Europe, tous ayant été infectés de façon congénitale. Le développement des rétinochoroïdites ainsi que la récurrence des lésions se produit plus tôt chez les enfants du Brésil. Leurs lésions sont aussi plus fréquemment multiples et de taille plus importante conduisant à un déficit visuel plus probable que chez les enfants européens (Gilbert et al., 2008).
Les problèmes oculaires ne touchent pas seulement les personnes ayant été infectées de façon congénitale mais aussi une large part des personnes ayant acquis l'infection après la naissance (Holland, 2003). Là encore, la fréquence des toxoplasmoses oculaires semble être plus importante en Amérique du Sud que nulle part ailleurs. Par exemple, une enquête populationnelle dans les années 70 à Walkersville, communauté rurale du Maryland aux USA, a montré que seulement 0,6% des 842 participants âgés de plus de 13 ans avaient une atteinte

BY-NC-ND

oculaire par *T. gondii* (Smith and Ganley 1972). A titre de comparaison, à Erechim dans l'état du Rio Grande do Sul au sud du Brésil, 21,3% des 1042 participants âgés de plus de 13 ans avaient une toxoplasmose oculaire, soit une prévalence 35,5 fois plus importante qu'à Walkersville, Maryland, USA (Glasner et al., 1992). En Colombie, pays qui compte 41 million d'habitants, on estime qu'environ 1 million de personnes ont des cicatrices rétiniennes de toxoplasmose oculaire et qu'au moins 200 000 ont une cécité unilatérale due à cette infection (de-la-Torre et al., 2009).

Les formes sévères de toxoplasmoses avec atteintes viscérales sont rares chez l'immunocompétent (Carme et al., 2002). Cependant leur fréquence est assez élevée en Guyane Française où plusieurs cas ont été reportés (Carme et al., 2002, 2009; Demar et al., 2007; Groh et al., 2012). Environ un tiers des patients atteints par ces formes de toxoplasmose présente une détresse respiratoire et nécessite d'être pris en charge dans une unité de soins intensifs. Les toxoplasmoses sévères en Guyane Française ont également été associées à des génotypes du parasite très divergents par rapport aux génotypes des souches isolées dans le reste du monde (Ajzenberg et al., 2004; Carme et al., 2002; Demar et al., 2007). La plupart des patients rapportent des activités en lien avec la forêt, telles que l'ingestion d'eau de surface et la consommation de viande de gibier mal cuite qui sont des sources d'infection par des souches de *T. gondii* appartenant au cycle sauvage de la forêt amazonienne. La forêt amazonienne abrite plusieurs espèces de félidés sauvages, qui en l'absence des chats domestiques sont la source des oocystes présents dans l'environnement (Carme et al., 2009). Le fait que les formes sévères de toxoplasmoses sont plus fréquemment rencontrées en Amérique du Sud où la diversité génétique de *T. gondii* est la plus élevée, a amené à suggérer que les différences génétiques pourraient être à l'origine des différences cliniques notamment pour les toxoplasmoses oculaires (Gilbert et al., 2008; Khan et al., 2006a). Les souches atypiques tendraient à provoquer des dommages plus importants que le type II pour les toxoplasmoses congénitales et les personnes immunocompétentes (Ajzenberg, 2012; Delhaes et al., 2010). Dans le cas des toxoplasmoses congénitales, l'association entre génotype et sévérité de la maladie chez les enfants infectés reste controversée, bien que l'analyse de plusieurs études semblent indiquer que le type I et les génotypes atypiques soient plus associés aux formes cliniques symptomatiques que les types II et III (Rico-Torres et al., 2016). Une étude portant sur 86 isolats de patients avec une toxoplasmose congénitale a montré cependant que le type II pouvait aussi causer des formes sévères. Cependant, les types I et les génotypes atypiques moins nombreux n'ont pas été détectés dans le groupe de patients asymptomatiques ou avec des formes bénignes (Ajzenberg et al., 2002). Le stade de la grossesse auquel se produit l'infection par *T. gondii* apparait comme le facteur le plus important dans le pronostic clinique. La transmission au fœtus est plus rare lorsque la mère s'est infectée en début de grossesse, cependant lorsque le fœtus est atteint, les complications cliniques sont plus graves que lors d'une infection en fin de grossesse (Ajzenberg et al., 2002; Rico-Torres et al., 2016). Aux Etats-Unis, le sérotypage de 183 cas de toxoplasmose congénitale a révélé que les sérotypes autres que le II (NE-II) étaient plus souvent présents chez les enfants nés prématurément ou avec des manifestations cliniques sévères. Cependant, tous les degrés de la maladie ont été trouvés pour tous les sérotypes, soulignant le fait que le génotype du parasite n'est pas suffisant pour prédire la sévérité de la maladie (McLeod et al., 2012).

Les souches de différents génotypes semblent induire des réponses immunitaires différentes. Par exemple, l'étude des données disponibles pour les toxoplasmoses oculaires en Colombie et en France a conduit à proposer deux mécanismes différents de réponse immunitaires en fonction de la souche infectante. Chez les patients français, la réponse cytokinique dans les

BY-NC-ND

humeurs aqueuses était de type TH1 avec production d' IFN-γ, IL-2, IL5 ainsi qu'une réponse Th17 avec augmentation de l'IL-17. Inversement, les patients colombiens avaient une réponse de type Th2 avec augmentation de l'IL-13 et une plus faible expression de cytokines pro-inflammatoires majeures comme l'IFN-γ, IL-2, IL15 et IL-17 qui étaient à peine détectables avec comme exception TNF-α et IL-6 qui étaient augmentées (de-la-Torre et al., 2013). Au total, il semblerait que les souches sud-américaines entraineraient une diminution de la réponse immunitaire protectrice conduisant à une multiplication plus importante du parasite dans la rétine. Des différences de capacité de migration et de franchissement de certaines barrières naturelles ainsi que des taux de croissance différents entre les souches expliqueraient aussi les différences de virulence (Saeij et al., 2005).

La susceptibilité de l'hôte fait aussi partie des hypothèses pouvant expliquer la variabilité de la maladie, impliquant l'existence de prédispositions génétiques. Par exemple, l'étude de deux cohortes de cas de toxoplasmoses congénitales a montré une association entre les atteintes oculaires et cérébrales et le polymorphisme du gène *ABCA* ainsi qu'une association entre le polymorphisme du gène *COL2A1* et les atteintes oculaires (Jamieson et al., 2008). L'importance des facteurs de l'hôte a été mise en évidence lors d'une épidémie de toxoplasmose sévère dans un village du Suriname près de la frontière avec la Guyane Française. Sur les 11 cas, 5 ont pu être génotypés. Une seule souche serait impliquée dans ces 5 cas pour lesquels les conséquences de la maladie n'ont pas été les mêmes, conduisant à la mort d'une femme de 56 ans et d'un fœtus après 22 semaines de gestation. Pour une même souche, les hôtes peuvent donc réagir de façon différente. Dans ce cas-là, l'âge peut être mis en cause, puisque les patients qui n'ont pas survécu sont le plus âgé et le plus jeune (Demar et al., 2007).

L'adaptation de l'Homme au parasite influencerait aussi les conséquences de la toxoplasmose. Cependant, l'Homme ne jouerait pas un rôle significatif dans l'évolution du parasite puisqu'il ne s'agit pas d'une proie de félidés, et que donc sa résistance à *T. gondii* n'aurait pas d'impact sur sa transmission (Müller and Howard, 2016). Les souches responsables des toxoplasmoses sévères en forêt Amazonienne proviendraient de la faune sauvage où le cycle sauvage du toxoplasme impliquerait des félidés sauvages et une grande diversité d'hôtes intermédiaires (Ajzenberg et al., 2007; Carme et al., 2002). L'Homme n'est pas un hôte habituel mais plutôt accidentel de ce cycle et il n'est donc pas adapté à la diversité des souches présentes (Ajzenberg et al., 2007). Une autre étude s'est intéressée au lien entre l'origine ethnique et la fréquence de réactivation des rétinochoroïdites toxoplasmiques, dans une région d'Argentine, Misiones, récemment peuplée par les immigrants Germaniques et Slaves au début du 20ème siècle. Elle semble indiquer que les patients slaves et germaniques auraient plus de risques de présenter une réactivation de toxoplasmose oculaire que les patients d'origine hispanique, soulevant le problème de l'adaptation aux souches d'Amérique du Sud en fonction de l'origine ethnique. Les Amérindiens, en contact avec les souches Sud-Américaines depuis plus de temps sembleraient plus résistants à la toxoplasmose induite par ces souches. Le métissage entre Européens et Amérindiens a principalement concerné les personnes d'origine hispanique, qui ont migré en Argentine à partir du 16ème siècle et a pu permettre la transmission des gènes de résistance à la toxoplasmose oculaire des Amérindiens. Au contraire les Européens d'origine slaves et germaniques ont migré plus tard et ont été moins concernés par le métissage (Rudzinski et al., 2016).

L'Amérique du Sud se caractérise donc par une prévalence de l'infection par *T. gondii* élevée mais également par une fréquence plus importante des formes sévères que les autres régions. La mesure de l'implication des facteurs du parasite et des facteurs de l'hôte n'est pas claire. Il semblerait que les mécanismes expliquant la variabilité des manifestations cliniques de la

BY-NC-ND

toxoplasmose impliquent l'interaction complexe de l'ensemble de ces facteurs. L'adaptation de l'Homme au parasite semble aussi jouer un rôle important.

I.3. Etat de l'art sur l'histoire évolutive de *T. gondii* et de ses hôtes définitifs

Les études sur la diversité génétique de *T. gondii* ont fait émerger une structure de population de plus en plus complexe. Ceci et la découverte de formes cliniques particulières, associées à des génotypes dits « atypiques » ainsi qu'à des régions plus reculées, sauvages, a conduit à s'interroger sur l'évolution de ce parasite. Les principaux points débattus sont l'origine des 3 types clonaux, l'origine géographique de *T. gondii* et les datations de ces évènements. Cette histoire évolutive reste controversée et plusieurs modèles ont été proposés. Ces modèles évoluent avec la prise en compte de nouvelles souches et l'utilisation de nouvelles méthodes d'analyse, notamment l'étude de génomes complets. Afin de comprendre les hypothèses émises dans certains de ces scenarii, une première partie fait le point sur l'évolution des félidés, qui ont sans doute joué une part importante dans l'évolution et la distribution du toxoplasme.

BY-NC-ND

I.3.1. Histoire évolutive des félidés

I.3.1.1. La phylogénie

La famille des félidés comprend 37 espèces. Certains aspects ont rendu difficile la reconstruction de leur histoire évolutive notamment la difficulté de trouver et de distinguer les fossiles, les évènements de spéciation rapides et récents et le problème des convergences. Cependant, les avancées en biologie moléculaire, particulièrement dans le domaine du séquençage ont permis d'obtenir une phylogénie robuste. Celle-ci est basée sur l'analyse de 30 gènes nucléaires échantillonnés à travers les 37 espèces. Sept espèces appartenant au sous-ordre des féliformes ont été utilisées comme groupe externe. Les datations des évènements de divergence ont été réalisées en utilisant une calibration à partir de 16 fossiles. Cette phylogénie confirme la présence de 8 lignées, déjà évoquées dans des études antérieures (Johnson and O'Brien, 1997; Pecon Slattery and O'Brien, 1998; Pecon-Slattery et al., 2004). Elle permet en plus de résoudre leurs relations phylogénétiques (Johnson, 2006) (Figure 7).

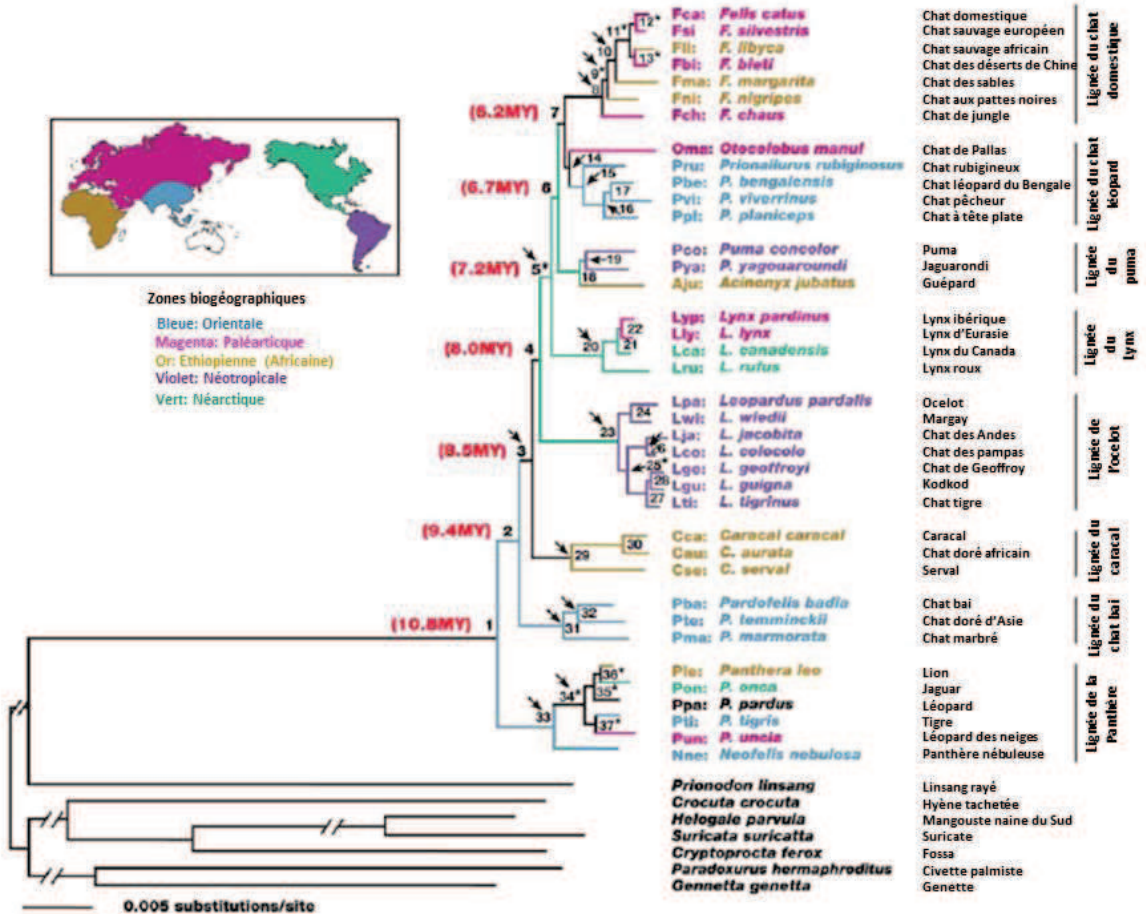


Figure 7 : Arbre phylogénétique des félidés, reconstruit avec la méthode de maximum de vraisemblance et à partir de 18853 bp d'ADN nucléaire. Données concaténées. Les noms scientifiques et les branches sont codées par couleur, chacune renvoyant à une zone zoogéographique (Orientale, Paléarctique, Ethiopienne, Africaine, Néotropicale et Néarctique)

Source : Adapté de Johnson (2006)

BY-NC-ND

Sur la base des fossiles, le genre *Pseudolarus* est considéré comme celui ayant pour descendance les chats modernes. Il aurait vécu en Asie il y a environ 11 Ma. A partir de cet ancêtre, la radiation des félidés modernes commence avec la divergence de la lignée des Panthères il y a environ 10,8 Ma, puis de la lignée du Chat Bai il y a environ 9,4 Ma. Cette dernière lignée comprend 3 espèces qui ont évolué et sont restées en Asie du Sud-Est. Au cours de l'histoire des félidés, six autres lignées vont apparaître : les lignées du Caracal, de l'Ocelot, du Lynx, du Puma, du Chat-Léopard, et du Chat Domestique (Figure 7). Les migrations ont joué un rôle important dans les différents évènements de spéciation.

BY-NC-ND

I.3.1.2. Les migrations

Les migrations des félidés sont favorisées par leur comportement qui les pousse avec l'âge à quitter leur territoire natal. L'augmentation de population se traduit donc par une expansion du territoire. Ce comportement à migrer est également renforcé par le suivi des populations de proies (O'Brien and Johnson, 2007). Ces migrations sont associées aux évènements géologiques et particulièrement à la variation du niveau des mers qui ouvre ou ferme des voies d'accès.

Aux alentours de 8 Ma, en raison du bas niveau des mers, l'Afrique et la Péninsule Arabique étaient reliées, permettant la migration de l'Asie vers l'Afrique d'un ancêtre de la lignée du caracal. Il y a 8,5 à 8 Ma, l'ancêtre de 5 lignées (ocelot, lynx, puma, chat-léopard et chat domestique) qui se différencieront plus tard, migre de l'Asie vers l'Amérique du Nord via le détroit de Béring. Le niveau de la mer remonte alors, séparant de nouveau les continents et favorisant la séparation des différentes lignées. La divergence de la lignée de l'ocelot a lieu vers 8 Ma, celle du Lynx vers 7,2 Ma et celle du Puma vers 6,7 Ma en Amérique du Nord (O'Brien and Johnson, 2007). Une troisième migration a lieu ensuite de l'Amérique du Nord vers l'Amérique du Sud vers 2,7 Ma. Elle a été favorisée par une nouvelle période glaciaire entraînant une baisse du niveau des océans et libérant une voie d'accès vers l'Amérique du Sud par l'isthme de Panama. Cette migration a notamment permis la diversification de la lignée de l'ocelot en Amérique du Sud. Les deux lignées du Chat Domestique et du Chat Léopard ont divergé des ancêtres Eurasiens qui sont restés en Asie ou qui ont migré en Amérique du Nord puis sont revenus en Asie entre 6,7 et 6,2 Ma (Figure 8).

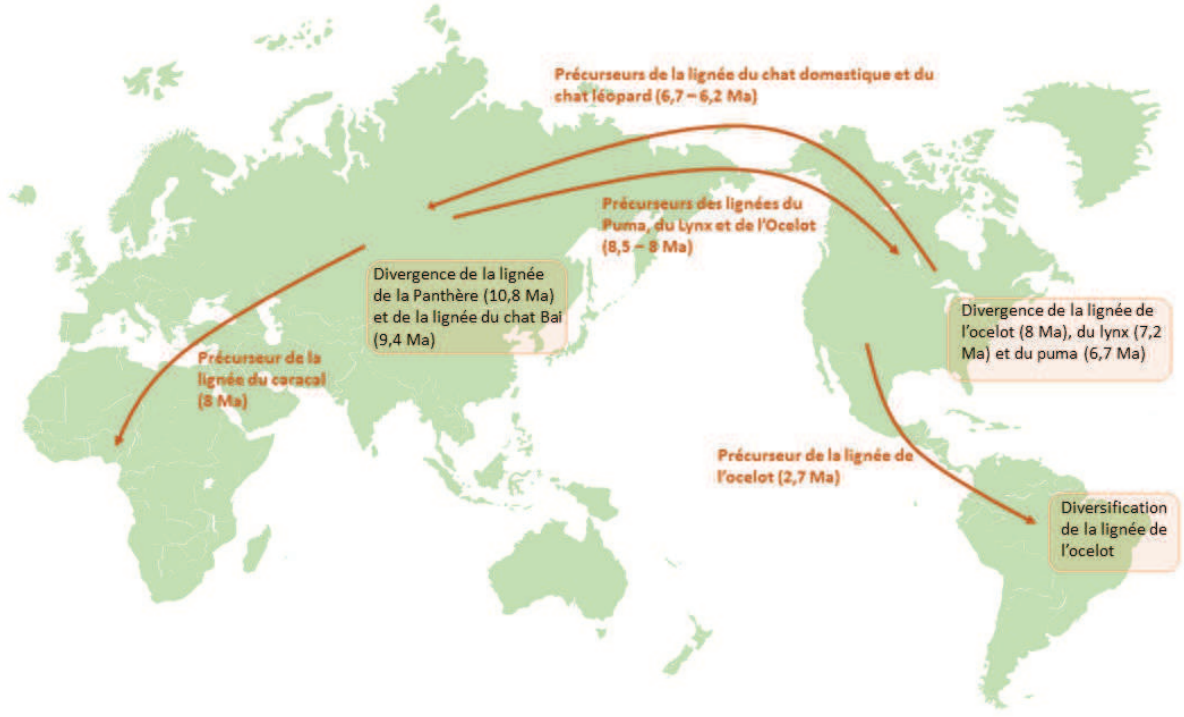


Figure 8 : Les migrations des Félidés expliquant la mise en place des huit lignées.

Source : Adapté de O'Brien and Johnson (2007)

BY-NC-ND

Ces quatre migrations permettent d'expliquer l'apparition des huit lignées. Cependant, parmi celles-ci, seules quatre se retrouvent uniquement dans les zones de leur établissement. Les autres lignées comprennent des espèces qui se retrouvent dans d'autres zones, ce qui implique des migrations supplémentaires. Dans la lignée du Puma, alors en Amérique du Nord, le guépard migre en Asie et en Afrique. Une autre migration de l'Amérique du Nord vers l'Eurasie, concernant la lignée du Lynx vers 1,6 à 1,2 Ma, va permettre l'établissement des lynx Ibérique et Eurasien. Des espèces de la lignée de la Panthère vont également migrer à partir de l'Asie : le jaguar et le lion en Amérique, et le lion et le léopard en Afrique (Figure 9). Enfin, plus tard 3 espèces de chats de la lignée du Chat Domestique migrent en Afrique (Johnson, 2006).
Après la période glaciaire, une extinction de masse a eu lieu, provoquant la disparition de 40 espèces de mammifères d'Amérique du Nord, notamment les lions d'Amérique mais la migration du guépard en Afrique, plus tôt dans l'histoire, permet d'éviter son extinction. Durant cette période, le puma trouve refuge en Amérique du Sud. Il ne reviendra en Amérique du Nord qu'entre 8000 et 10 000 ans.

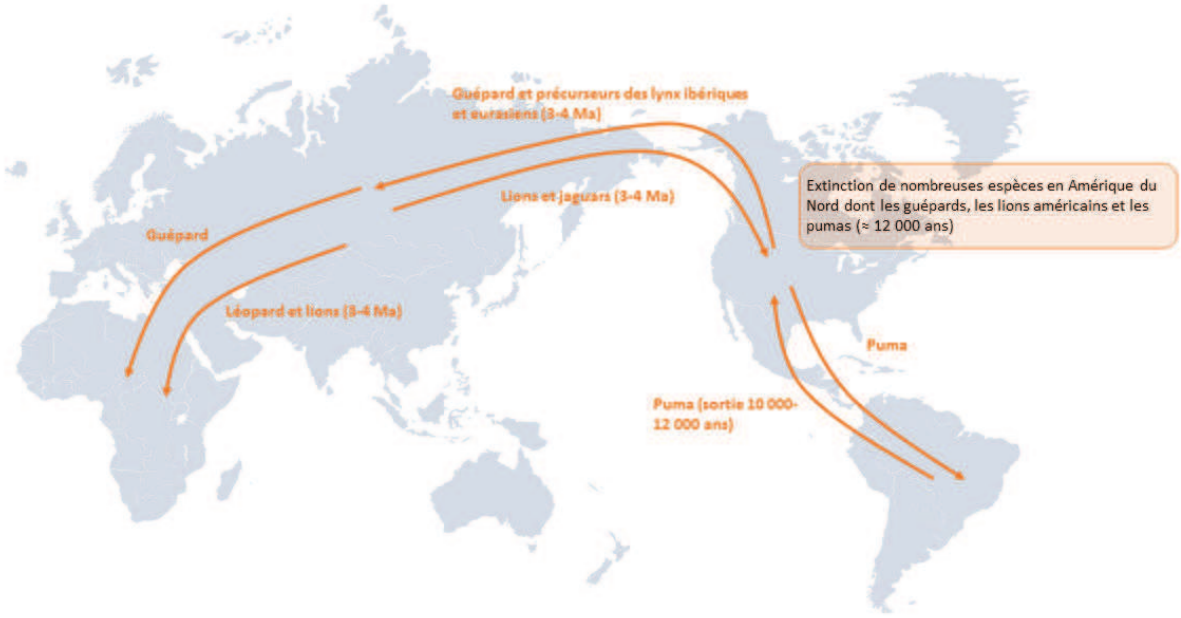


Figure 9 : Migrations des différentes espèces à l'intérieur des lignées, expliquant leur distribution actuelle.

Source : Adapté de O'Brien and Johnson (2007)

I.3.1.3. L'échange des faunes entre Amérique du Nord et Amérique du Sud

Au cours de cette histoire évolutive, le passage des félidés en Amérique du Sud vers 2,7 Ma fait partie d'un évènement biogéographique majeur : l'échange des faunes entre l'Amérique du Nord et l'Amérique du Sud ou grand échange interaméricain. Actuellement, la zone néotropicale américaine présente une richesse spécifique en mammifères importante, avec environ 1500 espèces, ce qui représente environ 30% de la diversité mondiale de mammifères (Carrillo et al., 2015). Cette situation est le résultat de la présence d'environnements variés ainsi que d'une longue évolution. L'Amérique du Sud a été isolée des autres continents durant

BY-NC-ND

plusieurs millions d'années lors du Cénozoïque, entraînant le développement d'une faune endémique (Carrillo et al., 2015). Cette situation cesse avec la connexion des deux Amériques par l'Isthme de Panama. Cet évènement s'est fait progressivement et il est difficile de le dater précisément. L'isolation de l'Amérique du Sud aurait cessé vers 10 à 7 Ma. Cependant, on considère que l'essentiel des migrations auraient eu lieu aux alentours de 3 Ma (Carrillo et al., 2015). C'est la date classiquement admise pour l'entrée des félidés en Amérique du Sud. Ce continent contenait plusieurs espèces de marsupiaux dont des carnivores mais aucun carnivore placentaire. Les marsupiaux présents n'étaient pas assez compétitifs face à des prédateurs comme les félidés, qui les ont progressivement remplacés. Ce succès des félidés en Amérique du Sud concerne particulièrement la lignée de l'ocelot qui a ainsi pu se diversifier rapidement (O'Brien and Johnson, 2007; O'Brien et al., 2008).

I.3.2. Origine des 3 types clonaux de *T. gondii*

Les premières études sur la diversité génétique de *T. gondii* ont mis en évidence une structure clonale composée principalement de 3 types clonaux. Une des hypothèses était alors que la structure de population de ce parasite correspondait à la théorie de la propagation clonale. Cette théorie repose sur le principe que même chez un organisme présentant une reproduction sexuée, les recombinaisons méiotiques contribuent peu voir pas du tout à la diversité trouvée. En général quelques lignées clonales prédominent, elles ont une origine ancienne et sont assez différentes. Une étude des souches appartenant aux 3 types clonaux a montré que la plupart des locus étudiés (15 locus) présentaient 2 classes d'allèles (Grigg and Suzuki, 2003). L'allèle partagé par au moins 2 types était soit identique soit légèrement différent (<0,5% de différence avec le consensus) et a été défini comme la classe allélique A tandis que l'autre allèle a été défini comme l'allèle E (Grigg et al., 2001). Le faible polymorphisme intra-lignée suggérait une origine récente des lignées clonales et il a donc été proposé que la clonalité observée chez ce parasite correspondait plus à une « clonalité épidémique », c'est-à-dire une expansion rapide et importante de quelques génotypes avantageux ayant réussi à coloniser la majorité des niches qu'à une clonalité. Avec ce scénario, alternatif de la théorie de la propagation clonale, les recombinaisons génétiques joueraient un rôle important dans la création de diversité, mais cette diversité serait « diluée » par l'expansion des types clonaux. Les premiers modèles proposés pour expliquer l'émergence de ces types clonaux impliquait le mélange de 2 lignées ancestrales, une seule recombinaison pouvant entraîner des différences de virulence importante (Grigg et al., 2001). Deux scénarios étaient alors proposés, notamment pour expliquer l'existence de variation dans la classe allélique A. Un faisait l'hypothèse d'un mélange ancien des deux pools de gènes, puis d'une accumulation de mutations par dérive génétique dans les souches, certaines de ces mutations ayant permis l'expansion des types clonaux, bien après le croisement (Figure 10.a) et l'autre faisait l'hypothèse d'un croisement tardif entre les deux pools de gènes, permettant l'accumulation de variation dans une classe allélique, juste avant l'expansion des types clonaux (Figure 10.b) (Grigg and Suzuki, 2003). Cependant le lien entre les types clonaux et les génotypes « atypiques » n'était pas clair. La majorité des souches « atypiques », bien que pouvant présenter des allèles entièrement nouveaux, étaient également considérés comme étant issues de la reproduction entre les 2 lignées ancestrales (Grigg et al., 2001). Cette dernière hypothèse est contredite dans une autre étude qui a montré que les mutations impliquées dans les souches exotiques n'affectaient pas les sites impliqués dans le système bi-allélique,

BY-NC-ND

démontrant que les souches atypiques auraient émergé avant les souches clonales et ne seraient donc pas issues du même croisement (Khan et al., 2007). L'apparition des souches clonales a été datée à environ 10 000 ans et celle des souches atypiques à environ 1 000 000 d'années (Khan et al., 2007; Su et al., 2003) (Figure 10.c). L'hypothèse expliquant cette émergence et cette prédominance des types clonaux était l'acquisition de la transmission par voie orale à travers les kystes (Su et al., 2003). Ces trois premiers modèles d'évolution de *T. gondii* ont été schématisés dans une étude qui a repris les estimations de la date du plus récent ancêtre commun (TMRCA) (Figure 10) (Morrison, 2005).

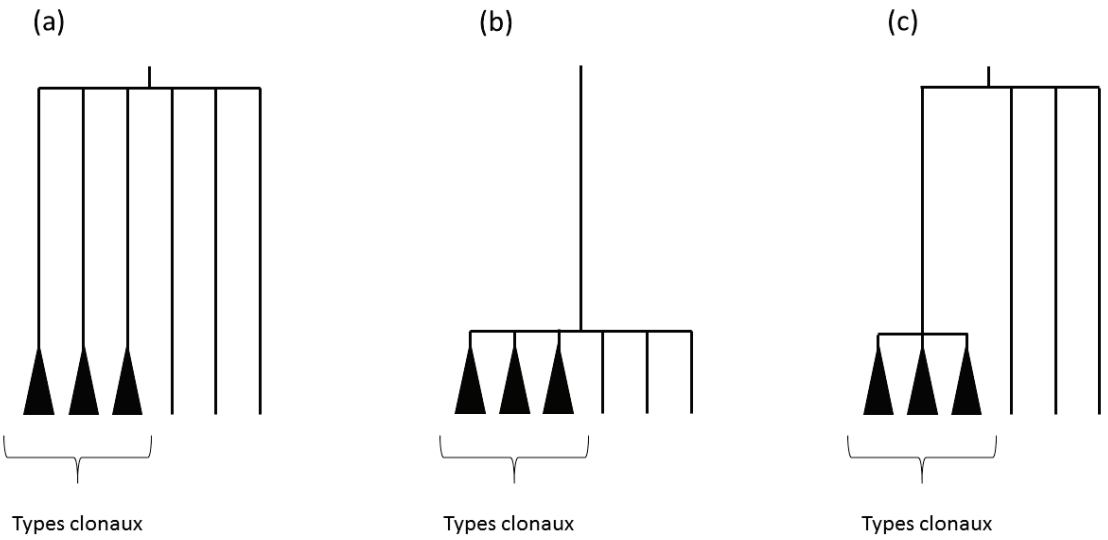


Figure 10 : Représentation schématique des 3 modèles d'évolution proposés pour *Toxoplasma gondii* : (a) et (b) à partir de Grigg et Suzuki (2003) et (c) à partir de Su et al. (2003). Les triangles représentent l'expansion des 3 types clonaux, les souches atypiques sont représentées par les autres lignes.

Source : Adapté de Morrison (2005)

Dans l'article de Morrison en 2005, les estimations réalisées, notamment le calcul des âges relatifs des lignées clonales et des lignées atypiques ne sont pas compatibles avec le modèle c et supportent plus vraisemblablement le modèle a. Le TMRCA de l'ensemble des lignées a été estimé à environ 150 000 ans. Il ne correspondrait pas à l'origine de l'espèce *T. gondii*, qui est plus ancienne aux environs de 11 Ma, mais plutôt à un évènement de goulot d'étranglement qui à partir d'une population de départ n'a laissé subsister que les lignées actuelles. Cette étude fait également la différence entre le TMRCA des lignées clonales daté de 150 000 ans et l'expansion de celles-ci daté de seulement 10 000 ans environ (Morrison, 2005), âge qui avait été attribué précédemment au plus récent ancêtre commun des lignées clonales (Su et al., 2003). Le TMRCA de 150 000 ans serait un âge minimum basé sur les souches prises en compte au moment de cette étude mais il est important de rappeler que l'inclusion d'autres souches pourrait donner une estimation beaucoup plus ancienne (Morrison, 2005).

A travers une analyse des SNPs sur l'ensemble du génome, un scénario plus précis de l'émergence des trois types clonaux a été proposé. D'après ce modèle, les trois types clonaux

BY-NC-ND

ne seraient pas issus d'un seul croisement entre 2 lignées ancestrales. Le type I serait issu du croisement entre un type II ancestral et une souche α, ou entre la descendance F1 du croisement précédent et une autre souche pouvant aussi être α, les calculs de probabilité favorisant le deuxième scénario avec 2 croisements. Le type III serait issu d'un croisement entre un type II ancestral et une souche β. Les souches hypothétiques de type II impliquées dans ces croisements seraient très similaires à l'ancêtre du type II moderne, tandis que les souches α et β seraient plus proches entre elles qu'avec le type II (Boyle et al., 2006). Ce modèle est présenté dans la figure 11 :

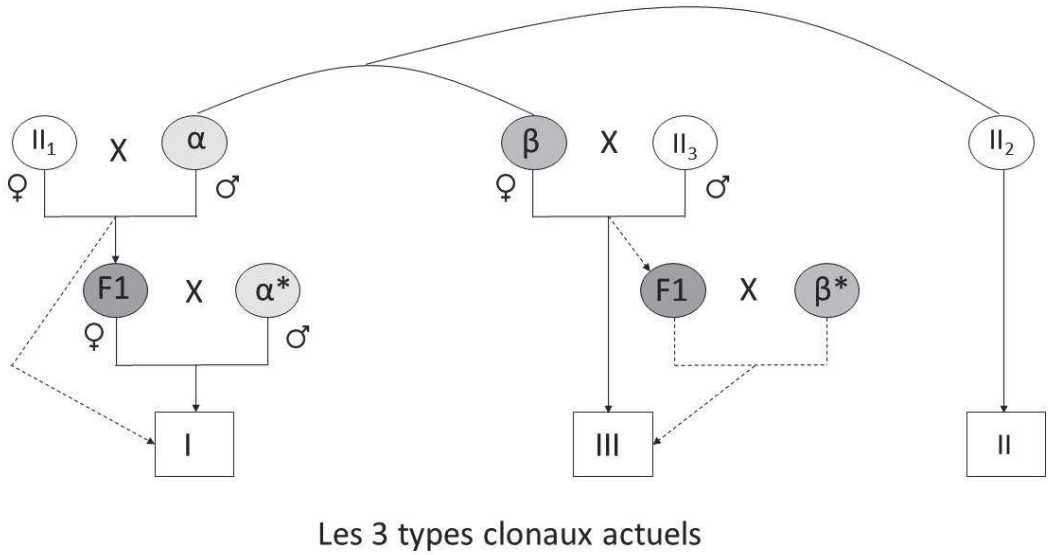


Figure 11 : Modèle proposé pour la création des 3 types clonaux de *Toxoplasma gondii*. En haut, les lignes connectant α, β et le type II représentent leur degré de relation. Les lignes pleines représentent les scénarios les plus probables, les lignes en guillemets, les scénarios alternatifs.

Source : Adapté de Boyle et al. (2006)

D'après le modèle proposé, les caractéristiques du génome des souches α et β ont pu être établies et une souche nommée P89 répondrait aux critères fixés pour la souche β. Elle descendrait donc de la souche parentale β qui aurait donné naissance au type III. Pour expliquer que seules quelques souches aient émergé de ces croisements pour devenir les types clonaux dominants, l'héritage d'une combinaison d'allèles leur procurant les traits les plus importants du type II et d'autres traits leur permettant de surpasser les souches parentales ainsi que celles issues du même croisement a été évoquée (Boyle et al., 2006).

BY-NC-ND

Ce modèle a été complété lors de la définition des 11 premiers haplogroupes. Il a alors été suggéré que la population ancestrale se composait de 4 lignées, et que la souche α, jusqu'alors inconnue s'apparenterait à l'haplogroupe 6. Une 4ème lignée proche de l'haplogroupe 4 aurait également contribué à l'émergence des autres haplogroupes (Khan et al., 2007) (Figure 12).

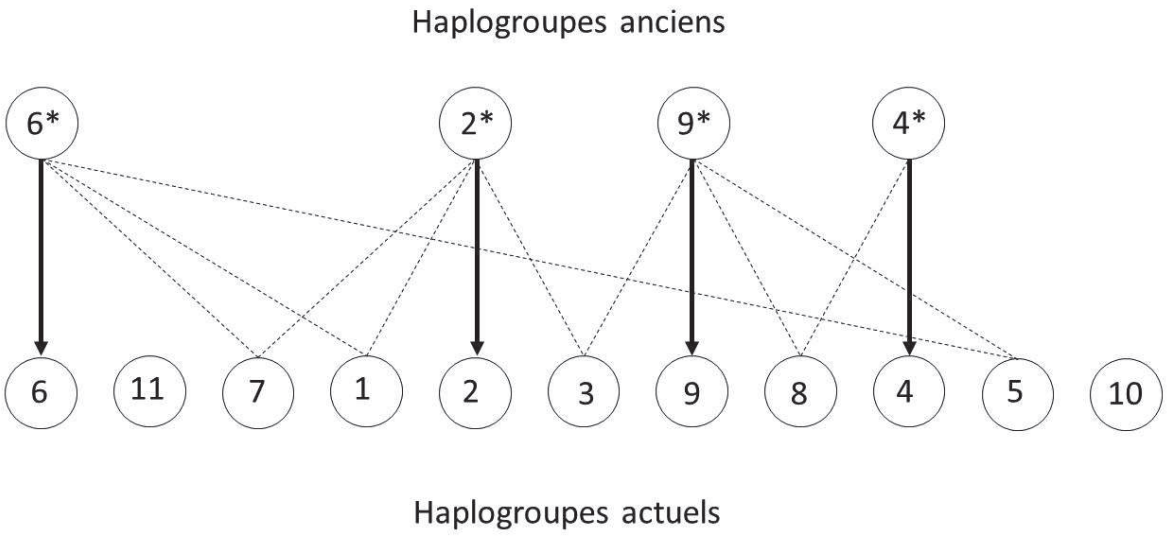


Figure 12 : Modèle pour la génération des 11 premiers haplogroupes à partir de groupes anciens ressemblant aux haplogroupes 2, 4, 6 et 9.

Source : Adapté de Khan et al (2007)

BY-NC-ND

De plus l'haplogroupe 12, commun dans la faune sauvage d'Amérique du Nord est également considéré comme une lignée clonale, et résulterait d'une recombinaison entre le type II ancestral et une souche γ, supposant qu'au moins 3 croisements avec le type II ancestral ont été nécessaires pour l'expansion des types clonaux (Khan et al., 2011a) (Figure 13).

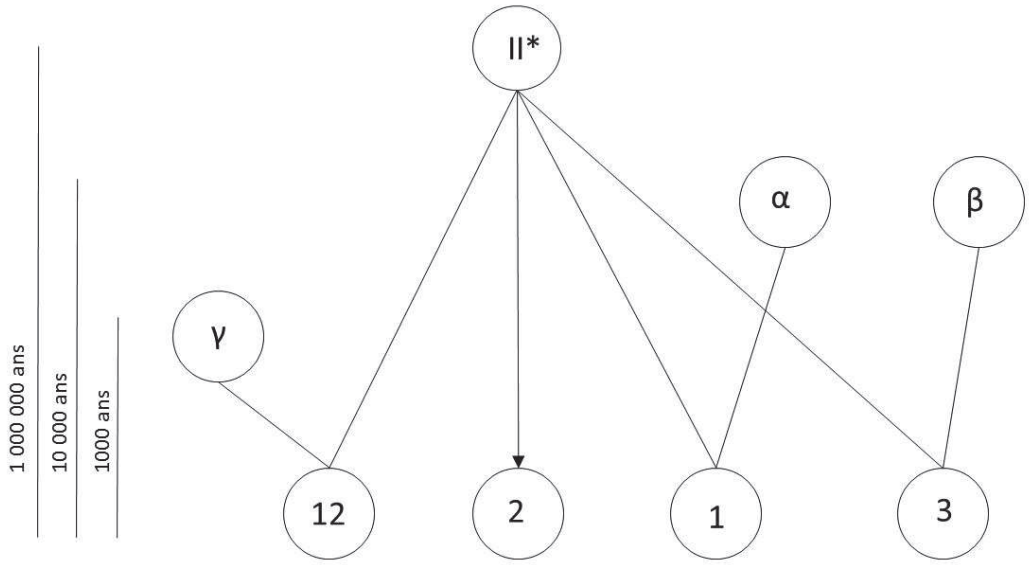


Figure 13 : Modèle proposé pour l'expansion des populations clonales de *Toxoplasma gondii* en Amérique du Nord et en Europe. Trois croisements séparés auraient eu lieu entre le type II ancestral et trois types parentaux différents nommés α, β, γ et auraient donné naissance aux types 1, 3 et 12 respectivement. Les dates indiquées sont issues d'études antérieures (Khan et al., 2007; Su et al., 2003).

Source : Adapté de Khan et al. (2011a)

Avec la découverte de nouveaux haplogroupes, au fur et à mesure de l'inclusion de nouvelles souches, 6 lignées ancestrales ont été proposées pour expliquer la structure de population de *T. gondii* (Khan et al., 2011b, 2011a). La vision des souches d'abord considérées comme « atypiques » ou « exotiques », car moins nombreuses dans les premières études et isolées principalement de régions reculées, change progressivement. Certains génotypes présents en Amérique du Sud montrent aussi, localement, une propagation clonale (Khan et al., 2007). Un polymorphisme bi-allélique est également mis en évidence dans les souches d'Amérique du Sud, bien qu'il soit exclusif de celui présent dans les souches clonales d'Amérique du Nord et d'Europe. Ces données suggèrent que ces deux groupes de souches ont une origine commune mais qu'ils ont été isolés il y a environ 1 000 000 d'années, et ont donc évolué indépendamment. Les haplogroupes montrant une propagation clonale, coalescent à 10 000 ans (Khan et al., 2007). La mise en évidence d'un très faible polymorphisme dans le chromosome Ia, aussi bien dans les souches clonales d'Amérique du Nord et d'Europe que dans beaucoup de souches d'Amérique du Sud, suggère qu'il contiendrait une combinaison d'allèles avantageuse pour les génotypes (Boyle et al., 2006; Khan et al., 2006b, 2007, 2011b) Cette version du chromosome Ia serait apparue récemment il y a environ 10 000 ans (Khan et al., 2011b). Sa présence dans plusieurs génomes qui semblent partager un TMRCA à 1 000 000 d'années suggère donc une diffusion rapide, associée à l'émergence des lignées

BY-NC-ND

clonales (Khan et al., 2007, 2011b). Cependant ces différents modèles ont été remis en question lors d'une analyse de SNPs sur l'ensemble du génome. En effet, cette étude a analysé plusieurs croisements proposés précédemment et a montré que des portions du génome de certaines souches n'étaient pas en accord avec ces croisements. Ainsi par exemple, 30% à 40% du génome des types I et III et de l'haplogroupe 8 ne ressemblent pas aux génomes des souches considérées comme parentales et la souche GUY-KOE, appartenant à l'haplogroupe 5, ne semble présenter aucune région vraiment similaire avec les haplogroupes 9 et 6 (Minot et al., 2012), présentés comme proches de ses lignées ancestrales (Khan et al., 2007). Ces résultats montrent que les groupes considérés comme parentaux ne sont pas les bons ou qu'il y a eu plus d'un croisement. Le génome de *T. gondii* a été découpé en haploblocs qui sont des régions génomiques presque identiques entre certaines souches ce qui suppose que ces haploblocs partagent un ancêtre commun récent. Cette analyse a montré que certaines souches appartenant au même haplogroupe ne présentaient aucun ancêtre commun récent pour les différents haploblocs identifiés, suggérant qu'elles sont en fait très divergentes. Cette étude met en évidence une histoire évolutive complexe durant laquelle certaines régions génomiques ont été échangées et se sont répandues dans la population alors que d'autres régions génomiques n'ont pas connu cette expansion (Minot et al., 2012).

L'ensemble des données sur la diversité génétique de *T. gondii* et les différents modèles proposés pour l'émergence des types clonaux a conduit à l'élaboration de plusieurs scénarios pour tenter d'expliquer où était apparu le parasite et comment il s'était disséminé.

I.3.3. Débat sur l'origine géographique de *T. gondii*

En 2006, une étude basée principalement sur des microsatellites a calculé la diversité génétique de différentes populations en utilisant des souches isolées à partir de poulets dans différents endroits du monde. Au total 275 isolats et 7 locus ont été pris en compte. Les diversités les plus élevées ont été trouvées en Amérique du Sud. Une représentation en réseau des haplotypes a montré une séparation entre les souches d'Eurasie et celles d'Amérique du Sud mais aussi une relation étroite entre certains génotypes présents dans plusieurs régions géographiques suggérant une radiation récente d'un seul génotype présentant une bonne capacité de migration. Un modèle statistique bayésien permettant de regrouper les souches sans a priori sur leur origine géographique a également montré qu'elles formaient 4 populations, 2 confinées en Amérique du Sud et Centrale, 1 présente en Europe, Asie, Afrique et Amérique du Nord mais apparemment absente d'Amérique du Sud et 1 cosmopolite. Une analyse basée sur les Fst mesurant la divergence entre chaque population et une population « ancestrale » théorique a permis de montrer que les populations d'Amérique du Sud présentaient des Fst plus petits, suggérant qu'elles avaient moins divergé de la population « ancestrale » commune que les autres (Lehmann et al., 2006). Ces données ont amené à conclure que le lieu d'origine de *T. gondii* était l'Amérique du Sud, où seuls les félidés sauvages en faible abondance pouvaient servir d'hôtes définitifs au parasite, avant l'introduction du chat domestique au 16ème siècle. Ce scénario propose au moins 2 migrations à partir de l'Amérique du Sud. Une première vers l'Eurasie, un évènement rare mais vers un lieu où il y a une abondance de chats domestiques, aurait permis l'émergence de la population présente en Europe, Asie, Afrique, Amérique du Nord mais pas en Amérique du Sud. Le trafic maritime a pu être à l'origine de cette migration puis de la distribution des souches formant cette population aux autres continents. Au 18ème et 19ème siècle, le transport de produits

BY-NC-ND

agricoles d'Amérique du Sud et Centrale a pu permettre quelques introductions, ayant pu conduire à l'émergence de la population globalement répandue. Il a été proposé que l'isolation géographique de l'Amérique du Sud ait pu conduire à la division de 2 lignées ancestrales d'allèles. La plus faible fréquence de recombinants en Amérique du Nord et Eurasie pourrait alors s'expliquer par un temps plus court pour produire ces recombinants par rapport à l'Amérique du Sud qui serait l'origine de ce parasite, présentant d'ailleurs un taux d'autoreproduction élevé. Dans cette étude les isolats utilisés proviennent de poulets. Les poulets sont considérés comme un bon indicateur de la contamination de l'environnement puisqu'ils se nourrissent au sol. De plus, l'intérêt de s'intéresser à cet hôte était de prendre en compte des infections asymptomatiques. En effet lorsque des cas humains sont pris en considération, il y a souvent le biais que seules les infections symptomatiques soient présentes dans l'échantillon. L'échantillonnage de cette étude était assez important avec notamment quelques souches isolées en Afrique et dans le sud de l'Asie, deux régions souvent sous échantillonnées. Cependant, l'origine de *T. gondii* en Amérique du Sud est principalement basée sur le fait que la diversité y soit plus importante qu'ailleurs, ce qui est insuffisant en l'absence d'approche phylogénétique ou spatiale pour retracer l'histoire évolutive du parasite. D'ailleurs, cette hypothèse a été contredite dans une autre étude, qui suppose plutôt que les souches d'Amérique du Sud proviendraient d'Amérique du Nord et qu'elles auraient été introduites en Amérique du Sud avec leurs hôtes définitifs (Khan et al., 2007). Les félidés ont en effet connu une diversification lors de leur entrée en Amérique du Sud. Cette étude utilise le séquençage d'intron et ne fait pas d'hypothèse sur l'origine en elle-même. Mais elle fait l'hypothèse qu'une séparation a eu lieu entre les souches d'Amérique du Nord et d'Amérique du Sud il y 1 000 000 d'années lors de l'entrée de *T. gondii* en Amérique du Sud, ce qui contredit le scénario de Lehmann et al. (2006). Le problème de cette étude est qu'elle n'utilise que 46 souches principalement en provenance d'Amérique du Nord et du Sud. L'Europe est faiblement représentée et l'Asie et l'Afrique ne rentrent pas dans l'analyse.

I.4. Objectifs de la thèse

La thématique de l'origine des pathogènes est large. Elle peut amener à plusieurs questions et peut concerner plusieurs échelles temporelles et spatiales. On peut par exemple s'intéresser à l'origine d'un pathogène dans une localité particulière, par exemple dans le cas d'une épidémie. On recherche alors la source d'une maladie. Les problématiques abordées peuvent cependant impliquer des échelles spatiales et temporelles beaucoup plus grandes. On peut alors chercher à identifier l'ancêtre ou l'origine géographique d'un pathogène, le nombre et le lieu d'éventuels transferts inter-espèces, les espèces réservoirs ou encore à reconstruire les routes d'invasion du parasite. Ce type de question est peu accessible par des approches d'épidémiologie classique et fait donc appel à des méthodes d'épidémiologie moléculaire qui reposent sur l'analyse de la diversité génétique, notamment à travers des approches phylogénétiques et spatiales. Enfin, ce type de questions peut concerner tous types de parasites à condition que la diversité présente soit suffisante pour être analysée et informative (Guégan and Choisy, 2009).

Plusieurs exemples peuvent être pris pour illustrer le type de questions qui peuvent se poser ainsi que la diversité des parasites concernés et des méthodes mises en œuvre. Une épidémie de choléra est apparue à Haïti en octobre 2010 alors que la maladie y était absente. Cette maladie est provoquée par une bactérie *Vibrio cholerae*. Deux hypothèses ont été émises pour expliquer son apparition dans cette île: les facteurs climatiques et la transmission humaine. Des méthodes d'épidémiologie classique et de comparaison génétique d'isolats ont été

BY-NC-ND

réalisées. Elles montrent une proximité phylogénétique entre les isolats d'Haïti et des isolats du Népal. Une épidémie de choléra sévissait dans ce pays durant la même période. Il a alors été suggéré que des soldats Népalais qui sont venus en aide à Haïti en 2010 après un tremblement de terre pour le maintien de l'ordre, auraient été en contact avec le choléra au Népal et l'auraient introduit de façon involontaire à Haïti (Frerichs et al., 2012). Le VIH (Virus de l'Immunodéficience Humaine) est l'agent du SIDA, décrit pour la première fois au début des années 1980. Il s'agit d'une maladie émergente, qui est devenue une pandémie mondiale en 25 ans. Deux types de VIH sont présents, le VIH-1 et le VIH-2. Le VIH-1 est impliqué dans l'épidémie mondiale, tandis que le VIH-2 a une aire de répartition plus limitée et se trouve plus particulièrement en Afrique de l'Ouest. Des reconstructions phylogénétiques ont permis de montrer que les deux types de VIH sont proches des VIS (Virus de l'Immunodéficience Simienne). En combinant les données phylogénétiques et les données sur l'aire de répartition des singes, il a été montré que les chimpanzés représentaient le réservoir originel des SIV ayant conduit au VIH-1 chez l'homme et que l'origine géographique du VIH-1 se situait en Afrique centrale de l'Ouest. Le VIH-2 aurait quant à lui comme origine un VIS infectant les mangabeys (Hahn et al., 2000; Sharp and Hahn, 2011). L'aspect des relations phylogénétiques entre ces deux virus a également permis de supposer que plusieurs transmissions indépendantes à l'Homme ont eu lieu. Les datations moléculaires indiquent que les différents VIH seraient apparus entre 1920 et 1945 (Faria et al., 2014; Lemey et al., 2003) D'autres pathogènes ont une origine plus ancienne. Par exemple la lèpre causée par une bactérie, *Mycobacterium leprae* qui a été décrite en 1873 est une des maladies les plus anciennes connues affectant l'Homme. Son génome a subit une évolution réductive avec une diminution de sa taille ainsi que l'accumulation de plusieurs pseudogènes. L'étude de différentes souches a montré que son génome était bien conservé et qu'il s'agissait d'un pathogène hautement clonal avec une fréquence de SNPs de 1 pour 28 kb (Monot et al., 2005). L'étude de 84 SNPs informatifs pour environ 400 échantillons provenant de 28 localités différentes a permis d'identifier 16 sous-profils SNP répartis dans 4 grands groupes (Monot et al., 2009). Une corrélation ayant été trouvée entre les profils SNP et l'origine géographique des patients, un scénario, impliquant une origine en Afrique de l'Est puis une dissémination via les migrations humaines, a été proposé afin d'expliquer l'existence des 4 grands profils SNPs de la façon la plus parcimonieuse possible, en tenant compte des données géographiques et historiques (Monot et al., 2005). Cette hypothèse a été confortée par des analyses supplémentaires notamment phylogénétiques (Monot et al., 2009). *Trypanosoma cruzi* est un protozoaire flagellé, faisant partie des kinétoplastidés, responsable de la maladie de Chagas. L'analyse phylogénétique des séquences d'ARNr 18S a montré une séparation claire entre *Trypanosoma cruzi* et *Trypanosoma brucei* responsable de la maladie du sommeil. *T. brucei* appartient à un clade comprenant exclusivement des trypanosomes d'origine Africaine, suggérant que son histoire s'est déroulée sur ce continent. Au contraire *T. cruzi* appartient à un clade qui contient des trypanosomes avec des origines en Amérique du Sud et en Australie. De plus, les trypanosomes de ce clade se trouvent à la fois dans des hôtes placentaires et marsupiaux. Ces données suggèrent donc que *T. brucei* serait apparu en Afrique après la séparation de ce continent avec l'Amérique du Sud qui a eue lieu il y a environ 80-100 Ma (Stevens et al., 1999). Au contraire, la maladie de Chagas est l'une des rares maladies infectieuses humaines à être originaire d'Amérique et non de l'Ancien Monde (Europe, Asie, Afrique) (Wolfe et al., 2007). Enfin, en ce qui concerne le groupe auquel appartient *T. gondii*, les Apicomplexes, une étude a montré à l'aide d'une analyse spatiale de la diversité génétique que *Plasmodium falciparum*, agent du paludisme, serait originaire d'Afrique sub-saharienne et que la diversité intra-populationnelle diminuait avec

BY-NC-ND

l'augmentation de la distance à l'origine. Un modèle génétique a ensuite permis de déterminer que ce pathogène aurait infecté les Hommes avant leur sortie d'Afrique il y a environ 50 000 – 60 000 ans et se serait disséminé dans le reste du monde via leurs migrations (Tanabe et al., 2010). Les exemples précédents montrent donc que plusieurs méthodes sont disponibles (analyse phylogénétique, analyse détaillée des SNPs, analyse spatiale de la diversité génétique, datation…) pour élucider l'origine d'une épidémie ou l'origine ancestrale de pathogènes émergents comme le VIH ou plus anciens (*Mycobacterium leprae, Trypanosoma cruzi, Plasmodium falciparum*). Elles nécessitent souvent d'être associées à d'autres informations telles que les origines géographiques des échantillons, l'aire de répartition des hôtes ou encore leurs migrations.

Ces études sur l'origine des pathogènes peuvent avoir diverses implications. Identifier la source d'une épidémie n'a pas toujours de conséquence directe sur la gestion initiale de la maladie mais elle permet toutefois d'identifier des facteurs de risques et de les prévenir en cas de situation similaire. Le cas de l'épidémie de choléra à Haïti peut par exemple conduire à des mesures de surveillance et de détection chez les contingents envoyés dans d'autres pays et où le risque d'introduction de nouveaux pathogènes peut être augmenté, notamment par l'affaiblissement des structures sanitaires provoqué par une catastrophe naturelle. L'identification de l'origine des VIH a permis de souligner l'importance de la pratique de la chasse et de la découpe de la viande de brousse dans le processus de transmission d'un virus animal à l'Homme. Cette découverte montre également l'importance de centrer une partie des recherches sur les VIS qui circulent en milieu naturel, notamment pour les reconnaître suffisamment tôt en cas de nouveaux transferts à l'Homme et d'éviter ainsi la circulation d'un nouveau virus non-détecté pendant plusieurs années (Guégan and Choisy, 2009). Plus généralement des facteurs de risque d'émergence de maladies infectieuses peuvent ainsi être mis en évidence, comme l'avancée de plus en plus marquée des populations humaines dans des régions inhabitées et sauvages et donc l'augmentation du contact avec des pathogènes nouveaux (Smith and Guégan, 2010). La détermination de l'origine de pathogènes plus anciens peut apporter des informations sur leurs mécanismes d'évolution et notamment sur le lien entre les degrés de coévolution et la susceptibilité plus ou moins importante des humains aux différents parasites. Il a ainsi été suggéré que la résistance des humains à la majorité des trypanosomes transmis par des mouches tsétsé en Afrique, pourrait être due à une coévolution plus longue entre l'homme et ces parasites sur ce continent alors que les trypanosomes d'Amérique ont majoritairement évolué sans contact avec les humains (Stevens et al., 1999). La connaissance de l'origine géographique d'un parasite permet également d'expliquer la distribution de sa diversité génétique et peut servir à la mise en place de moyens de lutte plus efficaces contre la maladie qui tiennent compte de la diversité. Ainsi la diversité génétique de *P. falciparum* s'est mise en place avec les migrations humaines et s'explique principalement par la distance à l'origine. Les campagnes de lutte contre la malaria ainsi que les variations dans la distribution géographique des vecteurs ne semblent pas avoir un effet significatif sur cette diversité (Tanabe et al., 2010). La diminution de la diversité intra-populationnelle observée pour des gènes de ménage a également été trouvée pour des antigènes reconnus par le système immunitaire de l'hôte et candidats pour l'élaboration de vaccins. De plus il a été montré qu'une large proportion de SNPs trouvés dans les populations non-Africaines se retrouvait aussi en Afrique supposant que ces SNPs sont stables dans le temps. Ces résultats suggèrent qu'un vaccin prenant en compte un petit nombre de génotypes d'antigènes sera plus efficace dans des populations présentant peu de diversité donc, plus éloignées de l'origine africaine, et qu'un vaccin prenant en compte les SNPs stables sera aussi plus efficace qu'un

BY-NC-ND

vaccin ne prenant en compte qu'un génotype. Les pressions de sélection exercées par le système immunitaire sur les antigènes ne semblent pas être le principal facteur affectant la distribution de la diversité. Cela diminue le risque que l'introduction d'un vaccin c'est-à-dire une pression de sélection, ne favorise la sélection de variants qui seraient capables de contourner le vaccin (Tanabe et al., 2013). Les applications en santé publique peuvent donc être nombreuses et plus ou moins directes.

L'objectif principal de cette thèse est de déterminer l'origine géographique des lignées actuelles de *T. gondii*. Savoir où est apparu l'ancêtre commun des souches actuelles permettrait sans doute de mieux comprendre l'évolution de ce parasite et la distribution géographique des différents génotypes. Ces résultats pourraient aussi avoir une implication dans la compréhension des différentes formes cliniques de la toxoplasmose, plus sévères en Amérique du Sud. Pour répondre à cette problématique nous avons choisi d'étudier une large collection de souches d'origine géographique variée mais regroupées dans des aires géographiques restreintes pour donner un sens à la notion de population. L'objectif principal de cette thèse était de déterminer l'origine géographique de l'ancêtre commun des lignées actuelles du toxoplasme et de dater son apparition. L'objectif secondaire était de déterminer les relations phylogénétiques entre les différentes souches de *T. gondii*.

La deuxième partie de ce travail présente cette étude.

BY-NC-ND

Chapitre II. Etude expérimentale

II.1. Le jeu de données

II.1.1. Sélection des souches

Au total, 168 souches ont été sélectionnées pour cette étude. Elles se répartissent sur 5 continents en 13 populations : 3 en Europe (Autriche, France et Portugal), 2 en Afrique (Gabon et Ethiopie), 2 en Asie (Chine et Turquie), 2 en Amérique du Nord (Minnesota et Pennsylvanie), 1 dans les Caraïbes et 3 en Amérique du Sud (Guyane, Brésil Sud-Est et Brésil Nord-Est). Les populations ont été définies de façon géographique, en prenant des souches spatialement proches. Les génotypes basés sur la méthode des microsatellites ont été pris en compte afin d'éviter les clones. Un minimum de 10 souches était nécessaire par population. (Annexe 2.1 et figure 14)

Les souches d'Europe, d'Afrique et d'Asie ont été isolées à partir d'animaux domestiques, à l'exception d'une souche de Turquie qui provient d'un cas humain de transmission congénitale. Les souches utilisées dans ce travail se répartissent de la manière suivante : 6 de poulets et 6 de cochons au Portugal (n=12), 12 de moutons en France (n=12), 10 de poulets en Autriche (n=10), 9 de chats et 1 d'un cas de toxoplasmose congénitale en Turquie (n=10), 13 de chats et 1 de cochon en Chine (n=14) , 6 de moutons et 7 de chèvres en Ethiopie (n=13), et 12 de poulets et de 2 de chèvres au Gabon (n=14). Les cochons du Portugal proviennent de petites exploitations en gestion semi-intensive dans la région de Vinhais et étaient destinés à l'abattoir. Les isolats de poulets de cette population sont issus d'une autre étude dans des petites fermes en zone rurale et ne proviennent donc pas tous de la même ferme. De même, les poulets de la population d'Autriche ne sont pas issus de la même ferme. Les souches de la population de Turquie ont été isolées à partir de chats errants provenant de quartiers différents d'Izmir. Le cas de toxoplasmose congénitale correspond à un cas sévère diagnostiqué à Ankara. Les souches de la population de Chine proviennent de différentes provinces du pays. Les moutons ainsi que les chèvres à partir desquels ont été isolés les souches d'Ethiopie ne font pas tous partie d'un même troupeau. L'ensemble des animaux se répartissait dans deux zones différentes : l'Est et l'Ouest de Shewa. Enfin en ce qui concerne le Gabon, les souches proviennent aussi de plusieurs localités : La lopé, le village de Dienga et trois zones urbaines, Makokou, Libreville et Franceville.

En Amérique, la situation est différente avec des souches à la fois d'animaux domestiques, d'animaux sauvages et de cas humains. En Amérique du Nord, la population du Minnesota est constituée de 10 souches isolées à partir de loups et de 2 à partir de cerf de Virginie (n=12), et la population de Pennsylvanie de 3 souches isolées à partir de renard, 1 d'oie, 1 de cerf et 7 de coyotes (n=12). En Amérique du Sud les souches sélectionnées sont : 1 souche isolée à partir d'un grison, 1 de jaguar, 1 de chien et 14 de cas de toxoplasmoses Amazoniennes pour la population du plateau des Guyanes (n=17), 9 de poulets, 1 de jaguarondi, 1 d'oncilla et 1 d'âne pour la population du Nord-Est du Brésil (n=12), 1 de mouton, 6 de chats et 8 de capybaras pour la population du Sud-Est du Brésil (n=15). La population des Caraïbes est constituée de souches isolées à partir de 3 cas de toxoplasmose congénitale, 4 à partir de cas de réactivation chez des patients immunodéprimés, 6 de chats et 1 de macaque (n=14). En ce qui concerne la population du Minnesota, les 10 souches isolées à partir de loups proviennent d'une étude où une séroprévalence de 52,4% a été trouvée pour un total de 248 loups testés. Les deux cerfs étaient en gestation et ont été prélevés dans une zone de 200

BY-NC-ND

km^2 autour de la ville de Skime. Les souches de la population de Pennsylvanie sont issues d'une étude qui a porté sur un total de 367 animaux dont 160 étaient séropositifs pour la toxoplasmose (92/266 coyotes, 49/79 cerfs, 17/20 renards et 2/2 oies). Dans la population de Guyane il faut noter que la souche isolée à partir d'un chien et celle à partir d'un grison sont issues d'un milieu anthropisé alors que les souches humaines ont été isoléés chez des patients qui s'étaient infectés en mangeant du gibier chassé dans la forêt amazonienne ou en buvant de l'eau de rivière non traitée. En ce qui concerne la population du Nord-Est du Brésil, l'oncilla était un animal captif et le jaguarondi vivait dans un zoo depuis 8 ans et était nourri de viande pré-congelée. Pour la population du Sud-Est du Brésil, les chats ne proviennent pas tous de la même localité et il en est de même pour les capybaras.

L'origine et la localisation de l'ensemble des souches sont détaillées dans l'annexe 2.1.

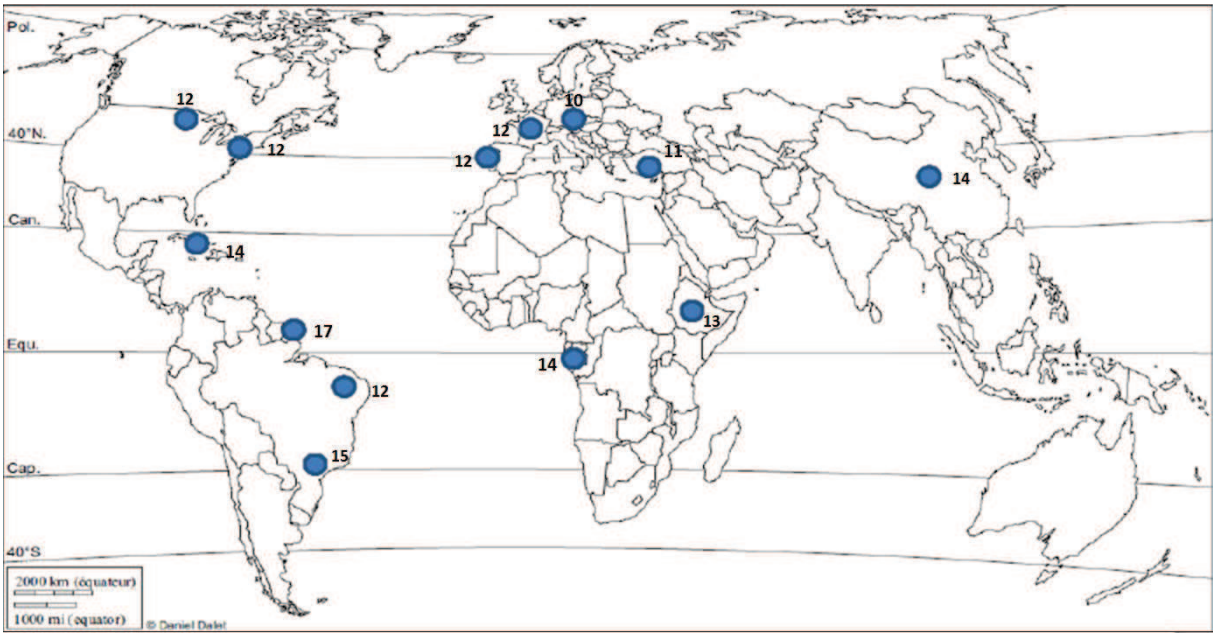


Figure 14 : Distribution géographique des populations et nombre de souches par population.

II.1.2. Sélection des marqueurs

Toxoplasma gondii étant un organisme présentant très peu de diversité dans certaines régions, le critère de sélection des marqueurs était le polymorphisme. Les fragments d'ADN devaient être suffisamment polymorphes pour différencier les souches et calculer des diversités génétiques. Pour cela, les régions d'ADN utilisées dans des études antérieures sur *Toxoplasma gondii* ont été recherchées sur la base de données GenBank (http://www.ncbi.nlm.nih.gov/genbank/) et ToxoDB (http://www.toxodb.org/toxo/). Les séquences de 30 fragments ont été téléchargées. A l'exception du marqueur correspondant à l'intron 1 du gène *TUB1* qui présente 6 séquences, tous les autres marqueurs avaient au minimum 16 séquences afin de permettre l'analyse. La nature et les fonctions de ces 30 fragments sont décrites ci-dessous :

- Les protéines de micronèmes : ces protéines sont sécrétées lors de l'invasion. Il s'agit en majorité d'adhésines c'est-à-dire de protéines de surface qui sont capables de lier des ligands spécifiques présents à la surface des cellules cibles (Dubremetz and Lebrun, 2012). MIC2 est une adhésine transmembranaire qui joue un rôle important dans la motilité du parasite lors de l'invasion. Cette adhésine fait partie d'un complexe

BY-NC-ND

qui permet de mettre en relation le glidéosome, moteur moléculaire, avec les récepteurs de la cellule hôte. Le déplacement de ce complexe vers le pôle postérieur permet le mouvement du parasite vers l'avant (Dubremetz and Lebrun, 2012; Frénal and Soldati-Favre, 2013).

- Les protéines de rhoptries : lors de l'invasion, les rhoptries libèrent leur contenu dans la cellule hôte à la suite des micronèmes. Le rôle de toutes ces protéines n'a pas été élucidé mais certaines sont impliquées dans la modification des voies de signalisation de la réponse immunitaire et ont été principalement étudiées dans le modèle murin. ROP18 est une sérine/thréonine kinase active, qui se place sur la surface membranaire de la vacuole parasitophore. Elle phosphoryle les IRGs (immunity-related GTPases) des souris, ce qui les empêche de provoquer la vésiculation de la vacuole parasitophore et donc à terme la destruction du parasite (Hunter and Sibley, 2012).

  ROP16 est une protéine qui cible le noyau de la cellule hôte et qui induit une modification de la transcription, en modulant l'activité des facteurs de transcription STAT3 et STAT6. Cette activité conduit à une diminution de l'inflammation et une réduction de la pathologie conduisant à une amélioration de la survie du parasite (Hunter and Sibley, 2012).
- Les protéines de granules denses : les granules denses sont des organites de stockage et de sécrétion de protéines, nommées les protéines GRAs. La sécrétion de ces protéines intervient vers la fin de l'invasion de la cellule hôte par le parasite lorsque la vacuole parasitophore est formée. Le manque d'homologie de ces protéines entre elles et avec d'autres protéines ayant une fonction connue a rendu difficile la détermination de leur fonction (Mercier and Cesbron-Delauw, 2015; Mercier et al., 2002). Cependant, il a été montré que ces protéines jouent un rôle dans la maturation de la vacuole parasitophore en un compartiment métaboliquement actif. Elles sont également importantes pour la transformation ultérieure de la vacuole en kyste (Mercier and Cesbron-Delauw, 2015). Parmi ces protéines GRA6, GRA7 et GRA3 sont associées aux différentes parties du système membranaire de la vacuole (Mercier and Cesbron-Delauw, 2015). Elles sont aussi impliquées dans d'autres rôles. Il a été montré par exemple que GRA6 régulait l'activation d'un facteur de transcription de l'hôte, le NFAT4, qui est un facteur nucléaire des cellules T activées, impliqué dans la virulence chez la souris (Ma et al., 2014). GRA7 a été trouvé dans le complexe protéique comprenant ROP18. Elle ciblerait également les IRGs et permettrait donc une virulence accrue du parasite, toujours chez la souris (Alaganan et al., 2014).
- Les antigènes de surface SAGs : il s'agit de protéines recouvrant la surface du parasite. Elles sont ancrées par du glycosylphosphatidylinositol (GPI). *SAG1*, *SAG2* et *SAG3* codent respectivement pour les protéines P30, P22 et P43 (Lekutis et al., 2001). Ces protéines se trouvent à la surface des tachyzoïtes (Fazaeli and Ebrahimzadeh, 2007; Lekutis et al., 2001). P43 (*SAG3*) se trouve également à la surface des bradyzoïtes (Lekutis et al., 2001). Ces protéines sont impliquées principalement dans l'attachement et l'invasion de la cellule-hôte et la modulation des défenses immunitaires de l'hôte (Fazaeli and Ebrahimzadeh, 2007; Lekutis et al., 2001; Wang and Yin, 2014).
- L'enzyme uracil phosphoribosyltransferase : codée par le gène *UPRT*, il s'agit d'une enzyme impliquée dans la voie de récupération des pyrimidines. Cette voie de synthèse des pyrimidines utilise des pyrimidines pré-formées originaires du parasite ou de la cellule hôte (Fox and Bzik, 2002; Iltzsch, 1993). Une de ses étapes consiste

BY-NC-ND

en la métabolisation de l'uracile en uridine 5'-monophosphate (UMP), précurseur de toutes les pyrimidines, par l'enzyme uracil phosphoribosyltransférase (Fox and Bzik, 2002).

- L'actine : le gène de l'actine code pour la protéine ACT1. Il a été montré que l'actine permettrait la ségrégation de l'apicoplaste parmi les cellules-filles lors de la réplication (Andenmatten et al., 2013). Elle fait partie du glidéosome et serait aussi requise lors de l'invasion de la cellule hôte pendant l'étape d'entrée du parasite (Drewry and Sibley, 2015). Ce gène est indispensable à la survie du parasite (Andenmatten et al., 2013).
- La tubuline : La tubuline est une protéine structurelle des microtubules. Ils sont formés d'hétérodimères de tubuline α et β, et participent à la formation de diverses structures, telles que le conoïde ou le flagelle des microgamètes. Ils jouent donc un rôle important dans l'invasion de la cellule hôte et la reproduction. Ils interviennent également dans la ségrégation des chromosomes lors de la réplication (Morrissette, 2015). *TUB1* et *TUB2* codent respectivement pour l'α-tubuline et la β-tubuline.
- La protéine codée par le gène *PK1* : Il s'agirait d'une sérine-thréonine kinase (Ng et al., 1997). Elle ferait partie des MAP-kinases qui permettent l'amplification des signaux dans le noyau afin d'initier des processus cellulaires liés par exemple à la croissance ou au développement (Schemarova, 2006).
- La protéine codée par le gène *HP* : Il s'agit d'une protéine hypothétique sans fonction connue.
- Les différents locus du chromosome Ia : le chromosome Ia fait environ 1,92 Mb et comprend environ 211 séquences codantes pour des protéines. Sa particularité est de présenter très peu de polymorphismes (Khan et al., 2006b). En effet une version similaire de ce chromosome est partagée non seulement par les types I, II et III mais aussi par beaucoup de souches originaires d'Amérique du Sud (Khan et al., 2007). Pour les 3 types clonaux, le pourcentage de polymorphisme de ce chromosome a été estimé à 0,03% alors qu'il a été estimé à 0,65% sur l'ensemble du génome (Boyle et al., 2006). Afin d'étudier ce chromosome pour plusieurs souches, 12 blocs d'environ 800 à 900 bp se répartissant sur le chromosome ont été sélectionnés pour être séquencés (Khan et al., 2006b, 2007, 2011b). Les locus 2, 5, 6, 7, 8, 12, 13, 14, 16, 19, 21 et 26 correspondent à ces 12 blocs.

Certains de ces fragments sont des gènes de ménage tels qu'*UPRT* et les gènes codant pour la tubuline et l'actine. D'autres sont impliqués dans l'invasion de la cellule hôte par le parasite et/ou reconnus par le système immunitaire de l'hôte. D'autres n'ont pas de fonction connue. Pour chaque locus, la souche de laboratoire RH, a été enlevée. Elle a été isolée en 1939 et cultivée depuis dans de nombreux laboratoires ce qui a entrainé l'apparition d'une variabilité génétique entre les différentes lignées de cette souche (Howe and Sibley, 1994). Les souches dont l'origine géographique était inconnue ont également été enlevées. Pour chaque locus, les séquences ont été alignées avec le logiciel MEGA v5.2.2 en utilisant l'algorithme MUSCLE. Certaines séquences ne s'alignant pas avec l'ensemble ont été considérées comme des erreurs et ont été éliminées. Lorsque plusieurs exemplaires d'une même souche étaient présents, seule une copie a été gardée. Le nombre de sites variables a été calculé pour chaque locus avec le logiciel R 3.0.2 en utilisant le package adegenet (Tableau 1). Afin de vérifier que les souches d'un même pays n'étaient pas identiques pour ces marqueurs et donc que ceux-ci étaient suffisamment informatifs pour calculer des diversités génétiques, les noms des souches ont été remplacés par celui de leur pays d'origine et un arbre non-raciné en

BY-NC-ND

Neighbor-joining a été construit pour chaque alignement à l'aide du package APE du logiciel R.

Parmi ces marqueurs, cinq ont été sélectionnés parmi les plus polymorphes : *GRA6, GRA7, SAG3, UPRT1* et *UPRT7* correspondant respectivement aux introns 1 et 7 du gène *UPRT*. Les fragments qui semblaient être les plus adaptés pour cette étude compte tenu du critère de sélection choisi, le polymorphisme génétique, étaient donc constitués de gènes codant pour des protéines de granules denses (*GRA6* et *GRA7*), un gène codant pour un antigène de surface (*SAG3*) et deux introns d'un même gène de ménage (*UPRT1* et *UPRT7*).

ROP18 n'a pas été sélectionné malgré son taux de polymorphisme élevé afin de pas biaiser les analyses phylogénétiques car il s'agit d'un facteur important de virulence soumis à une forte pression de sélection et dont l'histoire évolutive diffère de celle du reste du génome (Khan et al., 2009).

BY-NC-ND

Tableau 1 : Polymorphisme des différentes régions d'ADN de *Toxoplasma gondii* présentes dans GenBank.

Locus	Taille alignement	Nombre de séquences	S/L	P	N° d'accession GenBank
GRA6	212	145	18/212	85	AB235426-AB235435, AB471758, AF239284-AF239292, AJ635333-AJ635336, AY964058-AY964060, DQ187387, DQ512729, EF512218-EF512239, EF512241-EF512265, EF585704-EF585715, EU180619, EU180621, EU180623, EU180624, EU573185, JX044178-JX044180, JX044183-JX044189, JX044191, JX044192, JX044194-JX044196, JX044199, JX044201, JX044203-JX044211, JX044214-JX044225, JX044227-JX044234, JX044236, JX044237, JX044240-JX044243
SAG3	157	24	13/157	83	AF340228, AF340229, AY187280, EF585683-EF585694, JX218224, KC928250-KC928254, GT1 (ToxoDB), ME49(ToxoDB), VEG (ToxoDB)
ROP18	1673	45	137/1673	82	EF092842, GQ243205, GQ243206, GQ243208-GQ243216, JX045318-JX045320, JX045322-JX045329, JX045331-JX045338, JX045340-JX045345, JX045347-JX045354
UPRT intron 1	399	140	31/399	78	AY143142, AY143145, EU573186, JQ679758-JQ679781, JQ679783-JQ679895
GRA7	415	130	30/415	72	DQ459444-DQ459458, DQ465955, DQ473315, EF626965-EF626970, EF639857-EF639859, EU157141-EU157183, HQ852155-HQ852159, JX045574-JX045586, JX045588-JX045596, JX045600, JX045604-JX045615, JX045617-JX045632, JX045634-JX045638
UPRT intron 7	459	140	24/459	52	AY143148, AY143149, AY143151-AY143153, JQ679896-JQ679931, JQ679933-JQ680031
ROP16	2129	17	105/2129	49	AAYL02000151, GQ249081-GQ249096
GRA3	663	24	29/663	44	JX044153-JX044163, JX044165-JX044177

BY-NC-ND

Locus	Taille alignement	Nombre de séquences	S/L	P	N° d'accession GenBank
locus 5 Chr Ia	503	74	20/503	40	JX044476-JX044489, JX044491, JX044493-JX044543, JX044545-JX044552
SAG2	351	42	14/351	40	AF249696-AF249698, AF357577-AF357582, AY187279, EU258524-EU258530, EU258532, EU258533, EU650329, EU650330, JX045470, JX045473-JX045477, JX045479-JX045491, JX045493
Hp intron 2	470	137	16/470	35	JQ679621-JQ679757
MIC2 intron 3	372	75	11/372	30	JX044322-JX044324, JX044326-JX044383, JX044386-JX044399
Actin intron	432	25	12/432	28	AF247381, AF247382, JX045422-JX045443, JX045445
locus 21 Chr. Ia	721	73	19/721	27	JX045166-JX045187, JX045190-JX045223, JX045225-JX045241
SAG1	1011	78	26/1011	26	AY187278, GQ253073, GQ253074, GQ253086-GQ253098, JX045355, JX045357-JX045359, JX045361-JX045367, JX045369-JX045397, JX045400-JX045421
locus 13 Chr. Ia	717	74	19/717	25	JX044860-JX044925, JX044928-JX044935
locus 6 Chr Ia	457	73	11/457	25	JX044553-JX044579, JX044581-JX044607, JX044609, JX044611-JX044628
MIC2 intron 1	412	75	9/412	22	JX044245, JX044247-JX044268, JX044270-JX044321
PK1	459	16	10/459	22	EU258505-EU258518, EU650327, EU650328
locus 8 Chr. Ia	747	74	15/747	21	JX044708-JX044717, JX044719-JX044782

BY-NC-ND

Locus	Taille alignement	Nombre de séquences	S/L	P	N° d'accession GenBank
locus 2 Chr Ia	731	73	14/731	20	JX044401-JX044456, JX044458, JX044460-JX044475
locus 14 Chr. Ia	742	73	14/742	19	JX044936, JX044939-JX044953, JX044955-JX045011
locus 26 Chr. Ia	744	74	14/744	19	JX045242-JX045266, JX045268-JX045277, JX045279-JX045317
locus 12 Chr. Ia	631	73	11/631	18	JX044783, JX044785, JX044787-JX044819, JX044821-JX044858
TUB2 intron 1	447	77	6/447	14	AF249701-AF249703, AY143119, AY143121, AY143122, JX045496-JX045507, JX045509-JX045536, JX045538-JX045543, JX045545-JX045558, JX045561, JX045562, JX045564-JX045572
locus 7 Chr Ia	740	74	9/740	13	JX044629-JX044647, JX044649-JX044697, JX044699, JX044701-JX044705
Actin	1563	23	15/1563	10	JX045422-JX045443, JX045445
locus 19 Chr. Ia	746	74	4/746	6	JX045089, JX045090, JX045092-JX045126, JX045129-JX045165
TUB1 intron 1	501	6	2/501	4	AF249699, AF249700, AY143106, AY143107, AY143109, AY143110
locus 16 Chr.Ia	539	74	1/539	2	JX045013-JX045047, JX045050-JX045088

P: Sites polymorphes par kilobase
S: Nombre de sites polymorphes (inclus les sites polymorphes présentant aussi des gaps)
L: Longueur des séquences (avec les gaps)

BY-NC-ND

II.1.3. Amplification et séquençage

Les 5 fragments *GRA6, GRA7*, *SAG3, UPRT1* et *UPRT7* se répartissent sur 4 chromosomes différents (Tableau 2), ce qui permet d'éviter que ces marqueurs soient tous physiquement liés. Afin d'être sûr d'amplifier et de séquencer les régions d'ADN contenant le polymorphisme estimé dans l'analyse précédente, les amorces ont été choisies dans la bibliographie ou bien créées à partir du logiciel en ligne Primer3Web, de façon à ce qu'elles encadrent la région d'intérêt. Les amorces ainsi que la taille finale des séquences utilisées dans cette étude sont présentées dans le tableau 2.

Tableau 2 : Position, longueur et amorces des loci utilisés dans cette étude.

Locus	Chromosome	Position du fragment amplifié sur le chromosome	Taille* (bp)	Amorces	Source
GRA6	X	7 193 682 - 7 194 472	607	F: GTAGCGTGCTTGTTGGCGAC R: TACAAGACATAGAGTGCCCC	(Fazaeli et al., 2000)
GRA7	VIIa	2 604 391 - 2 605 143	677	F: TAAAAGGGATTACTGCGGCG R: CCCATATCCTACTGGCGGG	Cette étude
SAG3	XII	511 192 - 511 871	638	F: CCAATGGTCGAAGGCAAGAG R: CTCCGGAAAACACAGTTCGG	Cette étude
UPRT intron 1	XI	2 718 724 - 2 719 340	574	F: TCCAGGACATCATCACGAGG R: CCGCTTAGTGGCTTGATTCC	Cette étude
UPRT intron 7	XI	2 722 126 - 2 722 885	657	F: TGGTCGTCGTCACTTGTTA R: GCAGCCTCACAACTAAAACT	(Khan et al., 2007)

Chaque réaction d'amplification a été réalisée dans un volume total de 25 µL contenant 0,5 µL de Taq polymérase Herculase II Fusion (Agilent Technologies), 5 µL de tampon Herculase 5X, 0,63µL de dNTP à 10 mM, 0,63 µL de chaque amorce à 10 µM et 2 µL d'ADN. Les conditions d'amplification étaient 2 min à 95°C pour la dénaturation initiale suivie de 35 cycles avec une étape de dénaturation de 20 s à 95°C, une étape d'hybridation de 20 s à 60°C pour *GRA6*, 57°C pour *GRA7*, 58°C pour *SAG3*, 61°C pour *UPRT1* et 56°C pour *UPRT7*. L'étape d'élongation finale a été réalisée à 72°C durant 3 min. Ces conditions ont été réalisées avec un thermocycleur GeneAmpPCR System 2720 (Applied Biosystems).
Les réactions de PCR ont été contrôlées avec 5 µL de produit PCR lors d'une électrophorèse sur gel d'agarose à 2% à 100V pendant 30 min. L'ADN a été coloré avec du Midori Green (Dominique Dutscher). Les produits PCR ont ensuite été purifiés sur colonne en utilisant le kit

BY-NC-ND

Nucleospin Gel and PCR Clean-Up (Macherey-Nagel) en suivant les instructions du fabricant. Chaque produit PCR a ensuite été séquencé en utilisant le kit BigDye Terminator v1.1 Cycle Sequencing (Life technologies) et les mêmes amorces que pour la PCR. La méthode de séquençage utilisée est la méthode SANGER. Les réactions de séquençage ont été faites dans un volume total de 15 µL contenant 2,5 µL de BigDye, 2,5 µL de tampon BigDye 5X, 4 µL d'amorce et 2 µL de produit PCR purifié. Chaque fragment a été séquencé dans les deux sens avec les amorces sens et anti-sens. Les réactions de séquençage ont été purifiées par un système de filtration en utilisant Sephadex G-50 Superfine (GE Healthcare). Le séquençage a été réalisé à l'aide d'un séquenceur automatique (ABI PRISM 3130xl, Applied Biosystems). Les chromatogrammes ont été corrigés avec le logiciel CodonAligner v.5.1.5 DEMO. Les séquences sens et anti-sens ont été assemblées avec le programme en ligne emboss MEGAMERGER (http://www.bioinformatics.nl/cgi-bin/emboss/megamerger). Les alignements multiples ont ensuite réalisés avec le programme MUSCLE du logiciel MEGA v5.2.2.

Le choix des amorces a été fait de façon à encadrer la région d'intérêt sans tenir compte de la structure du gène. Les fragments amplifiés et séquencés dans ce travail comportent donc des régions de différentes natures et sont représentés dans la figure 10. Les séquences de *GRA7* sont les seules à ne présenter que de la région codante et à ne pas entièrement contenir la zone d'intérêt. Il manque en effet 18 nucléotides dans la partie 3'. Cela vient de la position limite de l'amorce en 3' et du découpage des extrémités lors du nettoyage des séquences. A l'extrémité 5', les séquences *GRA6* contiennent les 3 derniers nucléotides de la région non traduite, suivis de la région codante incomplète en 3'. Les séquences *SAG3* contiennent la région codante incomplète en 5', suivie par la région non traduite (UTR) incomplète en 3'. Les séquences *UPRT1* comprennent l'intron 1 incomplet suivi par l'exon 2 et l'intron 2 incomplet. Les séquences *UPRT7* sont constituées de l'intron 6 incomplet, suivi par l'exon 7, l'intron 7, l'exon 8 et la région non traduite incomplète en 3'.

BY-NC-ND

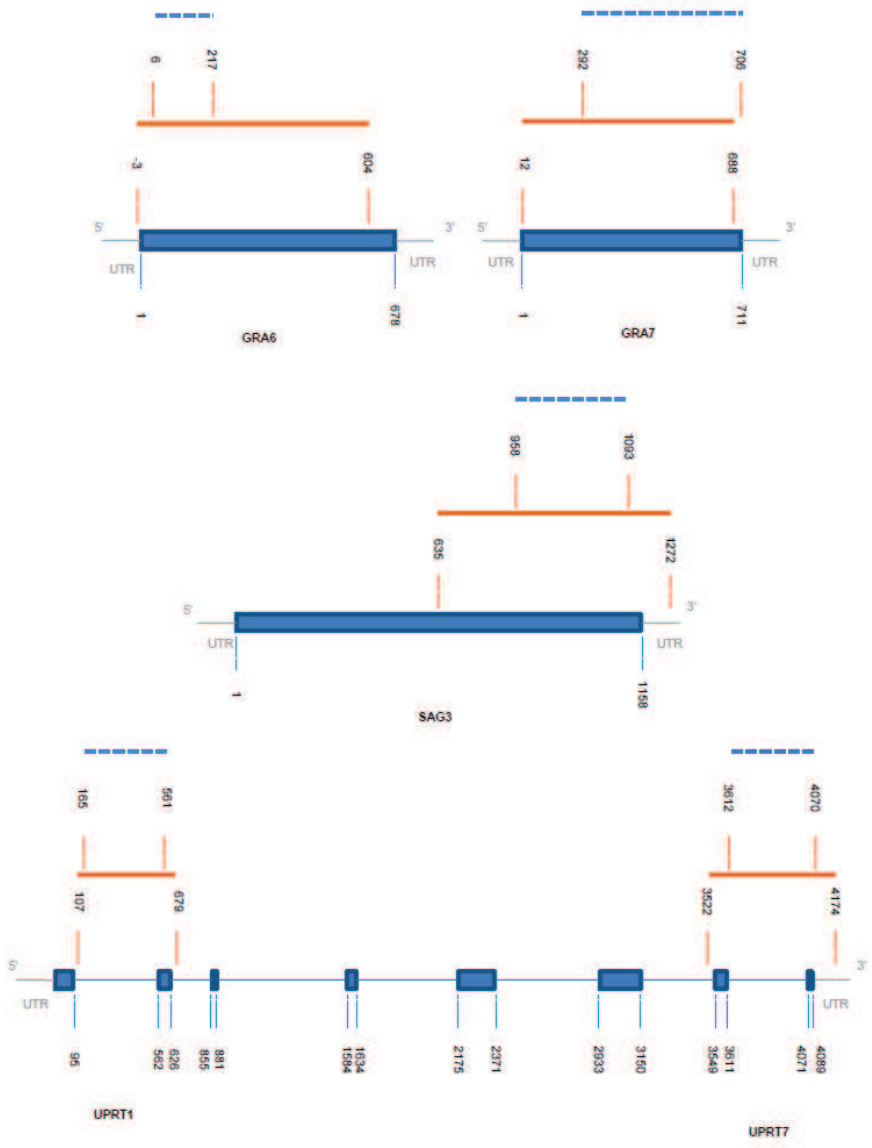


Figure 15 : Structure des gènes *GRA6*, *GRA7*, *SAG3*, *UPRT1* et *UPRT7* et localisation des fragments amplifiés dans cette étude (ligne continue orange), basée sur la souche de référence ME49. Les régions codantes sont représentées avec des rectangles pleins bleus, les introns en lignes continues bleues et les régions UTR en lignes continues grises. Pour chaque gène, l'emplacement des fragments issus de GenBank et utilisés dans l'étape de sélection des marqueurs (ligne pointillée bleue) et les zones séquencées dans cette étude (ligne orange) sont indiqués.

BY-NC-ND

II.2. Diversité génétique de *T. gondii*

La connaissance de la diversité génétique d'un parasite peut être utile pour mieux comprendre la maladie et par la suite la contrôler en mettant en place des stratégies d'intervention adaptées. En effet la diversité génétique d'un organisme lui permet de s'adapter à plusieurs environnements et donc de coloniser diverses niches écologiques. Elle lui permet également de subsister lors d'un changement, une diminution de la diversité génétique chez un organisme augmentant les risques d'extinction. Chez un pathogène, une telle diversité peut permettre de contourner les défenses immunitaires de l'hôte (Mita and Jombart, 2015). Il est donc important de connaître la diversité existante chez un organisme et notamment la diversité intra-populationnelle afin de comprendre ses variations géographiques. Cette partie présente les résultats de diversité trouvée au sein de chaque marqueur puis au sein de chaque population. Les différentes forces évolutives pouvant générer et influencer la diversité génétique sont d'abord rappelées.

II.2.1. Estimer la diversité génétique

La mutation est une force évolutive qui génère de la variation génétique. Elle se présente sous diverses formes telles que des mutations ponctuelles dues à des incorporations d'erreurs ou des insertions-délétions. Plusieurs facteurs (recombinaisons, migration, sélection, dérive génétique) influencent ensuite cette variation génétique, soit en modifiant les combinaisons des variations créées par la mutation, soit en agissant sur les fréquences des génotypes. Plusieurs formes de sélection existent. La sélection directionnelle permet à une mutation avantageuse d'augmenter en fréquence dans la population tandis que la sélection purifiante élimine les mutations délétères. La sélection balancée permet un taux de polymorphisme élevé à un locus en maintenant plusieurs allèles. Par exemple, pour un parasite, présenter divers types d'antigènes de surface l'aidera à échapper au système immunitaire et augmentera donc sa survie. Pour des locus différents, certains allèles peuvent être associés de façon préférentielle, il s'agit de la liaison génétique. Si un locus est en liaison génétique avec un autre qui est sous pression de sélection, il peut subir les mêmes effets par le phénomène d'auto-stop génétique. Un balayage sélectif se produit lorsqu'il y a une mutation très avantageuse à un locus. Sa fréquence augmente alors dans la population jusqu'à la fixation, entraînant aussi la fixation des allèles des loci qui lui sont liés. Des mutations plus faiblement avantageuses se trouvant sur d'autres locus peuvent ainsi disparaître de la population et des mutations faiblement délétères mais qui sont liées aux loci sous sélection peuvent se fixer. La sélection d'arrière-plan correspond au cas inverse c'est à dire à l'élimination de mutations délétères, pouvant entraîner la perte de mutations faiblement avantageuses à d'autres locus. Dans les deux cas, le polymorphisme diminue (Thomas et al., 2010). Tout le génome n'est donc pas soumis à la même pression de sélection et il est conseillé d'utiliser des régions considérées comme neutres ou presque neutres, telles que les substitutions synonymes pour estimer la diversité génétique. En génétique un site est considéré comme polymorphe s'il présente au minimum 2 allèles dans une même population. Ces allèles doivent avoir des fréquences suffisantes dans la population, ce qui est en général le cas lorsque l'allèle le plus commun a une fréquence inférieure à 95% (Mita and Jombart, 2015). Les recombinaisons, notamment au moment de la méiose peuvent créer de nouvelles combinaisons d'allèles et augmenter la diversité génétique. Cependant, ce processus ne crée de la diversité que si les deux souches qui subissent la recombinaison sont différentes. La migration est une force qui a tendance à augmenter la diversité. En effet, une réduction du flux génétique entre deux

BY-NC-ND

populations augmente le risque de consanguinité. Enfin la variation de taille des populations peut également influencer la diversité. La dérive génétique qui est responsable de la variation au hasard des fréquences alléliques peut conduire à la perte de certains allèles et donc diminuer la diversité. Les populations de petite taille sont plus sensibles à la dérive génétique.

Plusieurs indices peuvent être utilisés pour estimer la diversité génétique. Ceux utilisés dans ce travail sont présentés ci-dessous :

- le nombre de sites présentant 2 ou plus nucléotides, c'est-à-dire les SNPs, divisé par le nombre total de sites. Cette mesure est toutefois affectée par la taille de l'échantillon (Mita and Jombart, 2015).
- les distances entre séquences peuvent être calculées de différentes manières. Une distance évolutive est définie comme le nombre moyen de substitutions par site qui s'est produit depuis leur séparation. Ces distances peuvent être estimées en calculant le nombre de différences entre deux séquences. Cependant ce nombre correspond rarement à la distance évolutive puisque la comparaison de deux séquences ne permet pas d'avoir accès à toutes les substitutions qui se sont produites mais seulement à celles qui sont visibles. C'est pour prendre en compte ces évènements évolutifs que les distances entre les séquences sont calculées à l'aide de plusieurs modèles de substitutions nucléotidiques. Le modèle le plus simple, celui de Jukes et Cantor (JC), propose une évolution indépendante de chaque position du gène et une probabilité identique pour toutes les substitutions. A partir de ce modèle, plusieurs hypothèses ont été émises pour le complexifier. Le modèle K80 proposé par Kimura (1980) propose des probabilités différentes selon le type de substitution. Les transitions et les transversions n'ont pas la même probabilité de se produire.

  La distance de Hamming, elle, correspond au nombre de différences entre deux séquences, les allèles de chaque locus étant codé de façon binaire (1=présent, 0=absent) comme illustré ci-dessous :

	Locus 1		Locus 2		Locus 3	
	Allèle 1	Allèle 2	Allèle 1	Allèle 2	Allèle 1	Allèle 2
Seq 1	0	1	1	0	1	0
Seq 2	0	1	0	1	1	0

  Dans cet exemple, la distance de Hamming entre les deux séquences est de 2.

BY-NC-ND

- L'hétérozygotie attendue correspond à la probabilité que deux allèles tirés au hasard soient différents. Il s'agit donc d'une mesure applicable à un organisme haploïde comme *T. gondii*. Elle se calcule avec la formule suivante :

$$\frac{1}{K}\sum_{k=1}^{K}(1-\sum_{i=1}^{m(k)}f_i^2)$$

Où K est le nombre total de loci pris en compte, m(k) le nombre d'allèles au locus k et f_i la fréquence allélique de l'allèle i dans la population donnée.

- La richesse allélique a été calculée comme le nombre total d'allèles présents dans la population.
- La richesse haplotypique a été calculée comme le nombre d'haplotypes présents dans une population divisée par le nombre d'individus. Un haplotype a été défini comme une combinaison d'allèles pour les cinq marqueurs bien que ceux-ci se répartissent sur quatre chromosomes différents.

II.2.2. Calcul de la diversité génétique des populations de *T. gondii*

II.2.2.1. Détermination du polymorphisme par marqueur

Pour chaque population et chaque marqueur, le nombre de SNPs a été calculé en utilisant le package *adegenet* du logiciel R 3.0.2 et les indels ont été relevés.

Le nombre d'haplotypes par marqueur et leur fréquence ont été déterminés avec le package *pegas* du logiciel R 3.0.2. Les distances entre les séquences prises deux à deux ont été calculées, à l'aide du modèle d'évolution de séquences K80 avec le package *pegas* du logiciel R 3.0.2.

II.2.2.2. Détermination de la diversité génétique intra-populationnelle

Dans chaque population, la diversité génétique a été calculée à partir des sites extraits avec 4 indices après concaténation des 5 marqueurs : l'hétérozygotie attendue (Hs), les moyennes de la distance de Hamming, la richesse allélique et la diversité haplotypique.

BY-NC-ND

II.2.3. Résultats

II.2.3.1. Les sites polymorphes au sein des marqueurs

Le séquençage a permis de mettre en évidence deux types de polymorphisme : les SNPs et les indels (insertions-délétions). Les 5 fragments de gènes représentent un total de 3153 bp comportant 140 SNPs. Sur l'ensemble des populations, 26, 30, 27, 32 et 25 SNPs ont été identifiés dans *GRA6, GRA7, SAG3, UPRT1* et *UPRT7,* respectivement (Tableau 3). En tenant compte de la longueur des différents fragments, la diversité estimée à partir de ce nombre de SNPs est similaire pour *GRA6* (~0,043), *SAG3* (~0,042), et *GRA7* (~0,044). *UPRT1* a le plus fort taux de diversité (~0,055) et *UPRT7* le plus faible (0,038). Les indels présents étaient les suivants :

- Pour *GRA6*, une insertion de 3 nucléotides GAT en position 268-270 pour les souches GUY03, MIN04-MIN12, PEN07-PEN09
- Pour *GRA7*, une délétion de 3 nucléotides AAG en position 273-275 pour les souches CAR03, CHI02, CHI04-CHI14, MIN04-MIN09, MIN11, MIN12, PEN05, et PEN07-PEN09.
- Pour *UPRT1*, une insertion d'un nucléotide A en position 247 pour les souches BRS01-BRS02, BRS10, BRS12, GUY10, et GUY11
- Pour *UPRT7*, une insertion de 5 nucléotides en position 463-467 pour la souche GUY11

BY-NC-ND

Tableau 3 : Polymorphisme et composition des 13 populations.

Population	Nombre de souches	Nombre de SNPs					Total SNPs	Diversité Haplotypique*	Hs †	Distance de Hamming	Richesse allélique ‡
		GRA6	*GRA7*	*SAG3*	*UPRT1*	*UPRT7*					
Europe											
Autriche	10	0	0	1	0	0	1	0.200	0.001	0.400	141
France	12	9	14	15	0	0	38	0.333	0.043	13.303	178
Portugal	12	8	14	15	0	0	37	0.250	0.131	40.090	177
Asie											
Turquie	13	6	9	15	9	2	41	0.272	0.050	15.636	181
Chine	14	10	20	10	12	8	60	0.357	0.090	27.164	200
Afrique											
Ethiopie	11	8	14	15	0	0	37	0.230	0.067	20.615	177
Gabon	14	9	19	0	9	2	39	0.357	0.134	40.637	179

BY-NC-ND

Population	Nombre de souches	Nombre de SNPs					Total SNPs	Diversité Haplotypique*	Hs †	Distance de Hamming	Richesse allélique ‡
		GRA6	*GRA7*	*SAG3*	*UPRT1*	*UPRT7*					
Amérique du Nord											
Minnesota	12	16	15	15	0	1	47	0.583	0.088	26.939	187
Pennsylvanie	12	15	21	15	0	0	51	0.500	0.111	33.969	191
Amérique du Sud											
Plateau des Guyanes	17	19	15	18	18	20	90	1.000	0.133	39.588	231
Brésil Nord-Est	12	10	22	4	13	10	59	0.416	0.119	36.515	199
Brésil Sud-Est	15	13	22	2	18	10	65	0.666	0.166	49.866	205
Caraïbes	14	14	24	18	7	2	65	0.785	0.128	38.703	205
Total	168	26	30	27	32	25	140				

*: Diversité haplotypique = nombre d'haplotypes / nombre d'isolats
†: Hs = Hétérozygotie attendue
‡: Richesse allélique = Nombre d'allèles

BY-NC-ND

UPRT7 se distingue des autres marqueurs par un nombre élevé d'haplotypes. Il présente en effet 25 haplotypes tandis que *GRA6, GRA7, SAG3* et *UPRT1* en ont respectivement 15, 15, 14 et 16 (Figure 16). Le fragment *UPRT7* a ainsi le plus petit nombre de SNPs mais le plus grand nombre d'haplotypes. La distribution des haplotypes diffère entre *UPRT1* et *UPRT7* d'un côté et *GRA6, GRA7* et *SAG3* de l'autre. Pour *UPRT1* et *UPRT7*, un haplotype majoritaire présente une fréquence très élevée. L'observation des données montre qu'effectivement, pour ces deux fragments, l'haplotype majoritaire se trouve dans toutes les populations à l'exception de la Guyane et du Sud-Est du Brésil. En outre, toutes les souches d'Europe, d'Ethiopie et d'Amérique du Nord ont l'haplotype majoritaire d'*UPRT1* et *UPRT7* (seule 1 souche du Minnesota, MIN04, diffère au niveau d'un seul SNP de l'haplotype majoritaire d'*UPRT7*). La situation est différente pour *GRA6, GRA7* et *SAG3* pour lesquels les haplotypes majoritaires présentent une fréquence moins élevée (Figure 16).

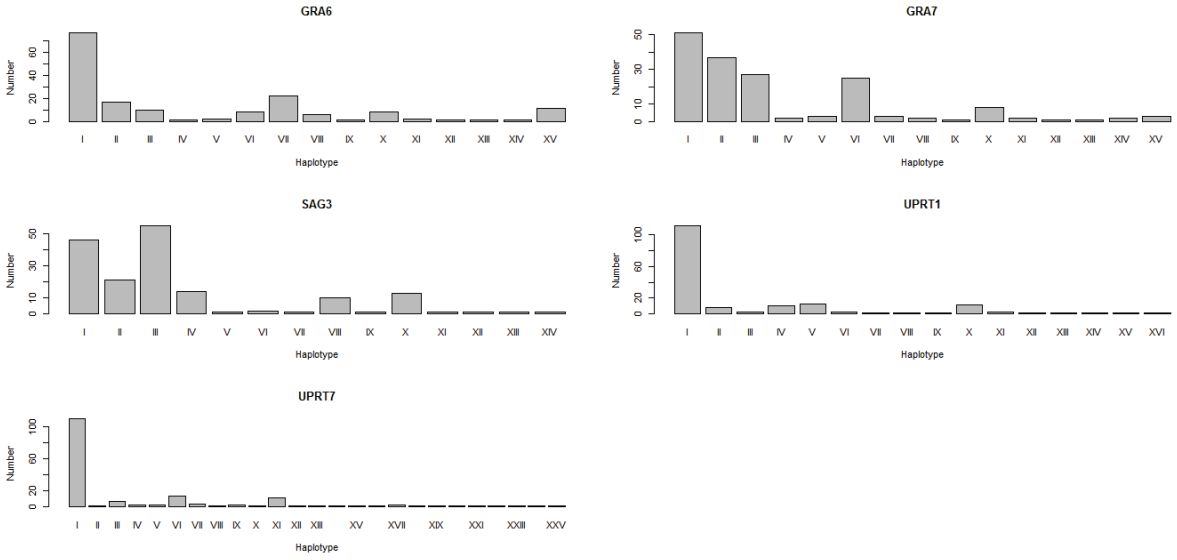


Figure 16 : Nombre et fréquences des haplotypes pour *GRA6*, *GRA7*, *SAG3*, *UPRT1* et *UPRT7*.

BY-NC-ND

Les distances génétiques parmi les haplotypes montrent le même contraste entre les marqueurs *GRA6, GRA7* et *SAG3* d'un côté et *UPRT1* et *UPRT7* de l'autre. Les distances K80 sont en effet plus dispersées pour les 3 premiers marqueurs (Figure 17)

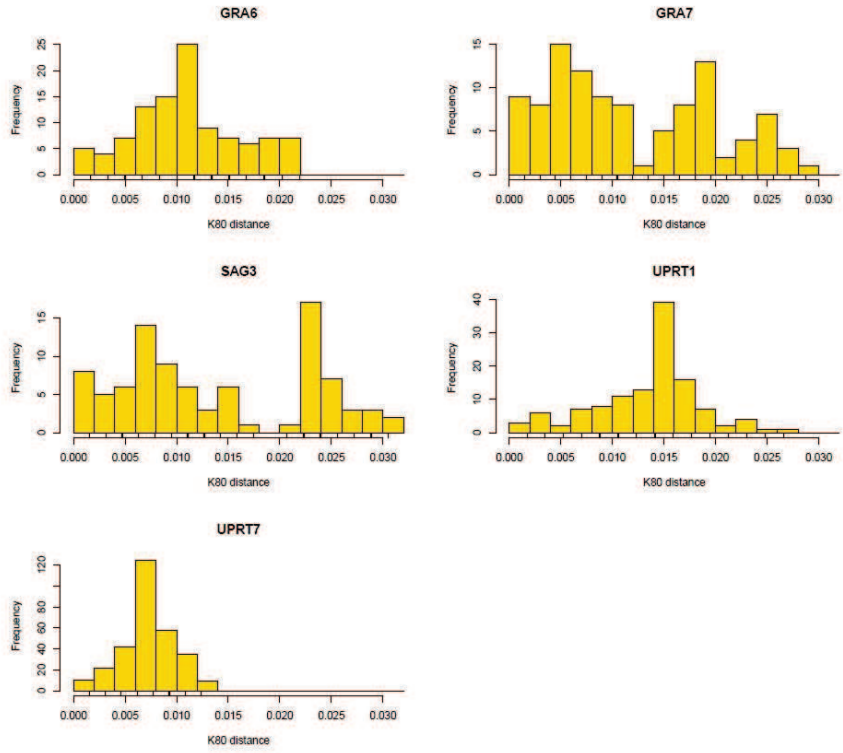


Figure 17 : Distributions des distances K80 deux à deux pour *GRA6, GRA7, SAG3, UPRT1* et *UPRT7.*

II.2.3.2. Les diversités génétiques des populations

Les souches d'Amérique du Sud présentent un nombre de SNPs variant entre 59 pour la population du Nord-Est du Brésil et 90 pour celle du plateau des Guyanes (Tableau 3). La population de Chine présente un polymorphisme assez semblable avec 60 SNPs. Le polymorphisme est moins élevé en Europe, en Turquie et en Afrique avec un nombre de SNPs qui varie entre 1 et 41, tandis que l'Amérique du Nord présente un polymorphisme intermédiaire, le nombre de SNPs variant entre 47 et 51 (Tableau 3).
La concaténation des données a permis d'identifier 60 haplotypes. Le nombre d'haplotypes varie de 4 en Europe à 32 en Amérique du Sud (Figure 18). Aucun haplotype n'est partagé entre l'Amérique du Sud et l'Amérique du Nord. Les Caraïbes qui occupent une position centrale entre ces deux régions partagent seulement 2 haplotypes avec l'Amérique du Nord, plus particulièrement avec la population de Pennsylvanie et 1 haplotype avec l'Amérique du Sud.

BY-NC-ND



Figure 18 : Nombre et fréquences des haplotypes présents en Asie, dans les Caraïbes, en Europe, en Afrique, en Amérique du Nord et en Amérique du Sud.

Les diversités haplotypiques sont plus élevées en Amérique du Sud, avec une valeur maximale (1) dans le plateau des Guyanes où tous les haplotypes présents sont différents. Le Brésil du Nord-Est présente une valeur un peu plus faible (0,416). Ces diversités sont plus faibles en Europe, Asie et Afrique avec des valeurs s'étalant de 0,2 à 0,357. L'Amérique du Nord présente des valeurs intermédiaires. La richesse allélique montre globalement les mêmes résultats. Cependant la moyenne des distances de Hamming et l'hétérozygotie attendue diffèrent. Avec ces indices, l'Amérique du Sud fait toujours partie des populations les plus diversifiées mais, c'est le Brésil du Sud-Est qui montre la diversité la plus élevée. De plus, le Portugal et le Gabon montrent des diversités similaires à celles de l'Amérique du Sud.

BY-NC-ND

II.2.4. Discussion

II.2.4.1. Polymorphisme et nature des marqueurs

Dans cette étude, il existe deux types de gènes avec apparemment des histoires évolutives différentes : des fragments d'un gène de ménage *(UPRT*) et des gènes codant pour des antigènes de granule dense et de surface (*GRA6, GRA7* et *SAG3*). Le nombre de SNPs ne semble pas différent entre ces deux types de marqueurs. En effet, les fragments d'*UPRT* présentent à la fois le plus petit (*UPRT7*) et le plus grand nombre de SNPs (*UPRT1*). Une autre étude a comparé des gènes de ménage et ceux codant pour des antigènes (différents de ceux de ce travail) et montré que la variation entre les marqueurs était très faible et n'était donc pas structurée selon ces deux groupes, bien que les introns de gènes de ménage apparaissent plus informatifs que les séquences codant pour des antigènes (Lehmann et al., 2000). En terme de SNPs il n'y a donc pas une grande différence entre les marqueurs, mais la distribution de leurs haplotypes n'est pas la même. L'analyse des haplotypes par marqueur révèle qu' *UPRT1* et *UPRT7* possèdent un haplotype fortement majoritaire. Ceci suggère que cet haplotype s'est plus largement répandu dans les populations que les autres. Les mutations présentes dans ces deux marqueurs se situent en grande majorité dans des régions non-codantes, a priori neutres du point de vue évolutif. Il est donc peu probable que ces haplotypes se soient répandus aussi largement parce qu'ils procuraient un avantage sélectif dans les populations concernées. Cependant, ces introns peuvent être associés à des exons présentant des mutations avantageuses et peuvent donc être soumis aux mêmes pressions de sélection par le phénomène d'auto-stop. Si cela a été le cas, ce balayage sélectif n'a pas concerné toutes les populations. En effet, les populations du Brésil du Sud-Est et plus particulièrement de Guyane présentent un nombre beaucoup plus élevé d'haplotypes pour ces deux marqueurs, avec par exemple 16 haplotypes pour *UPRT7* en Guyane sur un total de 17 souches. Les distributions des haplotypes et des distances génétiques entre séquences semblent plus équilibrées pour *GRA6*, *GRA7* et *SAG3*. Bien que ce ne soit pas le cas dans toutes les populations (celle d'Autriche est très homogène pour tous les marqueurs), le maintien de différents haplotypes à des fréquences plus équilibrées peut suggérer un avantage à maintenir une certaine diversité pour ces marqueurs. Il a été suggéré que les variations dans la protéine de GRA6 pourraient jouer un rôle important dans la survie du parasite (Fazaeli et al., 2000). De plus GRA6, GRA7 et SAG3 sont des antigènes, ils sont donc reconnus par le système immunitaire de l'hôte. Le maintien d'une certaine diversité dans ces gènes peut permettre au pathogène de contourner les défenses immunitaires de l'hôte (Mita and Jombart, 2015). Ces deux types de marqueurs semblent donc évoluer de manière différente. Il est donc possible que les modèles d'évolution à appliquer ne soient pas les mêmes et qu'il faille donc analyser les différents types de marqueurs de façon indépendante.

II.2.4.2. Diversité intra-populationnelle et divergence entre les indices de diversité génétique

La distribution des haplotypes peut permettre de faire des hypothèses sur les potentielles relations entre les différentes régions du globe. L'absence d'haplotypes partagés entre l'Amérique du Nord et l'Europe d'un côté et l'Amérique du Sud de l'autre suggère que peu d'échanges se sont produits entre l'Amérique du Sud et ces deux régions de l'hémisphère Nord. Au contraire, l'Amérique du Sud possède des haplotypes communs avec l'Afrique ce qui

BY-NC-ND

indiquerait la présence d'échanges entre ces deux parties du monde, comme cela a déjà été suggéré (Mercier et al., 2010).

Globalement une forte diversité est trouvée dans les populations d'Amérique du Sud, résultat cohérent avec plusieurs études (Ajzenberg et al., 2004; Khan et al., 2007; Lehmann et al., 2006; Su et al., 2012). Cependant, le détail des valeurs des quatre indices de diversité génétique montrent que certains sont divergents entre eux. La structure de population de *T. gondii* peut expliquer ces différences. Dans certaines régions, *T. gondii* se propage surtout de façon clonale, tandis que dans d'autres régions, les recombinaisons entre des souches différentes sont plus probables. Dans une population présentant seulement 2 haplotypes avec des fréquences similaires, la richesse haplotypique sera faible. L'hétérozygotie attendue sera elle plus élevée puisque la fréquence de chaque allèle sera proche ou égale à 50%. C'est le cas pour la population du Portugal pour de nombreux locus dans les marqueurs *GRA6*, *GRA7* et *SAG3* et pour la population du Gabon dans les marqueurs *GRA6*, *GRA7*, *UPRT1* et *UPRT7*. Ainsi la propagation clonale d'un faible nombre d'haplotypes assez différents et présentant une fréquence similaire dans la population augmente artificiellement l'estimation de la diversité génétique avec l'hétérozygotie attendue. La distance de Hamming semble également être augmentée par la présence d'haplotypes très différents.

Ces résultats montrent que tous les indices de diversité ne sont pas équivalents et qu'il est important de choisir le plus approprié à l'organisme étudié.

II.3. Phylogénie

II.3.1. Les phylogénies moléculaires

Les arbres phylogénétiques sont utilisés dans plusieurs branches de la biologie. Ils représentent les relations de parenté entre espèces mais ils peuvent aussi, entre autres, décrire les histoires des populations et fournir des indices sur les dynamiques évolutives et épidémiologiques des pathogènes. Les arbres phylogénétiques peuvent être générés à partir de deux grands types de données biologiques : les caractères discrets et les mesures de distances (Yang and Rannala, 2012).

II.3.1.1. Les méthodes basées sur les mesures de distances

La première étape des méthodes basées sur les distances est de calculer une matrice qui regroupe les distances évolutives entre chaque paire de séquences. C'est cette matrice qui est utilisée par la suite pour la reconstruction de l'arbre phylogénétique. Plusieurs méthodes ont été mises au point afin d'inférer un arbre à partir d'une matrice de distance dont :

- La méthode UPGMA (Unweighted Pair Group Method with Arithmetic mean) : elle procède par agglomérations successives de séquences ou groupes de séquences. Elle consiste à grouper les séquences qui présentent la plus petite distance dans la matrice, puis à évaluer la nouvelle distance entre le groupe nouvellement formé et le reste des taxons. Cette méthode aboutit à la construction de phénogrammes, c'est-à-dire d'arbres ultramétriques, obéissant à l'horloge moléculaire qui fait l'hypothèse que les taxons évoluent à la même vitesse (Thomas et al., 2010). Cependant, cette hypothèse est souvent contredite, et la méthode UPGMA n'est presque plus utilisée.
- La méthode Neighbor-Joining (NJ) (Saitou and Nei, 1987) procède par agglomération successive de paires de taxons à partir d'un arbre en étoile. Le regroupement de 2

BY-NC-ND

taxons génère un nouveau nœud qui les représente et les remplace dans la matrice (Gascuel, 1997) (Figure 19). Le choix des nœuds à agglomérer est fait suivant le principe du minimum d'évolution qui consiste à sélectionner l'arbre avec les plus petites longueurs de branches (Yang and Rannala, 2012). Cette méthode permet de prendre en compte des vitesses d'évolution différentes entre taxons et s'affranchit donc de l'hypothèse de l'horloge moléculaire (Thomas et al., 2010). Il s'agit de la méthode de distance la plus largement utilisée.

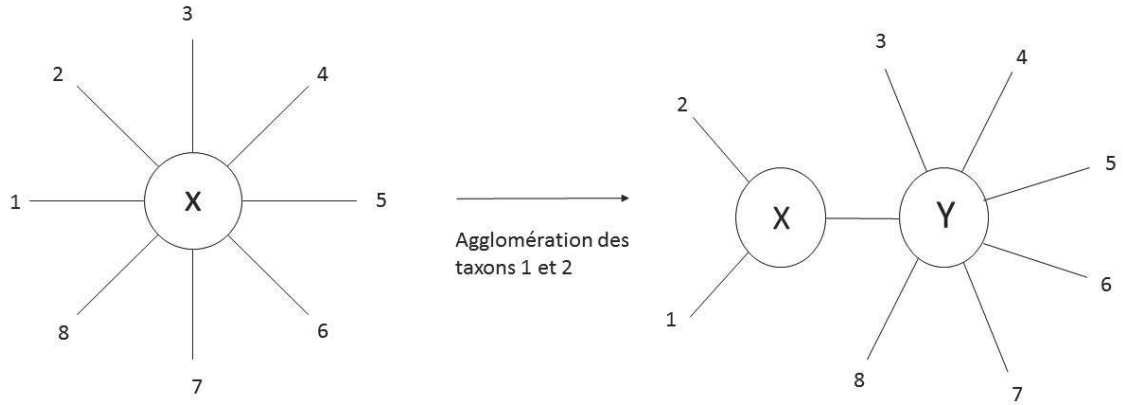


Figure 19 : Principe de l'agglomération du Neighbor-Joining. L'algorithme commence avec un arbre en étoile et agglomère une paire de taxons.

Source : Adapté de Yang and Rannala (2012)

- La méthode BioNJ est une amélioration de NJ qui prend en compte les variances et les covariances des distances évolutives. Il donne de meilleurs résultats que NJ, en particulier lorsque les taux de substitutions sont élevés et varient entre les lignées (Gascuel, 1997).
- La méthode FastMe réalise l'algorithme du minimum d'évolution de (Desper and Gascuel, 2002). Par rapport à NJ, il ne produit que des arbres avec des longueurs de branches positives et est plus rapide (Desper and Gascuel, 2004)

BY-NC-ND

II.3.1.2. Les méthodes basées sur les caractères discrets

Les méthodes d'inférence phylogénétique basées sur les caractères incluent la méthode du maximum de parcimonie et les approches probabilistes (Maximum de vraisemblance et approche bayésienne).

- Le principe de maximum de parcimonie repose sur une double optimisation. La première consiste à choisir pour chaque caractère l'arbre qui minimise les changements d'état. Cette approche nécessite de reconstruire les états de caractères aux nœuds de l'arbre. Seuls les caractères informatifs sont utilisés c'est-à-dire qu'au moins deux états doivent être présents chez les taxons utilisés et ces deux états doivent être présents dans au moins deux taxons. La deuxième optimisation consiste à choisir la topologie qui minimise le nombre de changements pour l'ensemble des caractères.
- Les méthodes probabilistes s'appuient sur des modèles explicites d'évolution de séquences et sur le concept de vraisemblance qui peut s'écrire selon la formule : $Lx = \Pr(X|H)$ qui représente la probabilité conditionnelle d'observer les données X selon l'hypothèse H. Afin de décrire l'évolution biologique des séquences de la façon la plus réaliste possible, les modèles peuvent incorporer plusieurs paramètres. Ils sont donc plus ou moins complexes. Les principaux modèles existants et les paramètres pouvant être pris en compte sont représentés dans la figure 20. Les méthodes probabilistes permettent d'estimer les valeurs de ces paramètres durant l'analyse (Delsuc and Douzery, 2004). Selon le modèle utilisé, les fréquences en A, C, G et T seront égales ou propres à chaque nucléotide. Les taux de substitutions peuvent également être égaux entre tous les types de substitutions, ou bien prendre en compte la différence entre transversions et transitions ou autoriser des taux différents pour tous les types. Une des hypothèses de base de certains modèles est que le taux de substitutions ne change pas entre les sites. L'incorporation d'une distribution gamma (Γ) qui décrit l'hétérogénéité entre les différents sites permet d'envisager des taux de substitutions différents entre les sites. Enfin, les modèles peuvent également prendre en compte une fraction de sites invariables (I). Il est important de choisir le modèle qui est le mieux adapté au jeu de données. La sélection de ce modèle peut s'effectuer selon le critère d'information d'Akaike ou AIC (Akaike Information Criterion) (Delsuc and Douzery, 2004).

BY-NC-ND

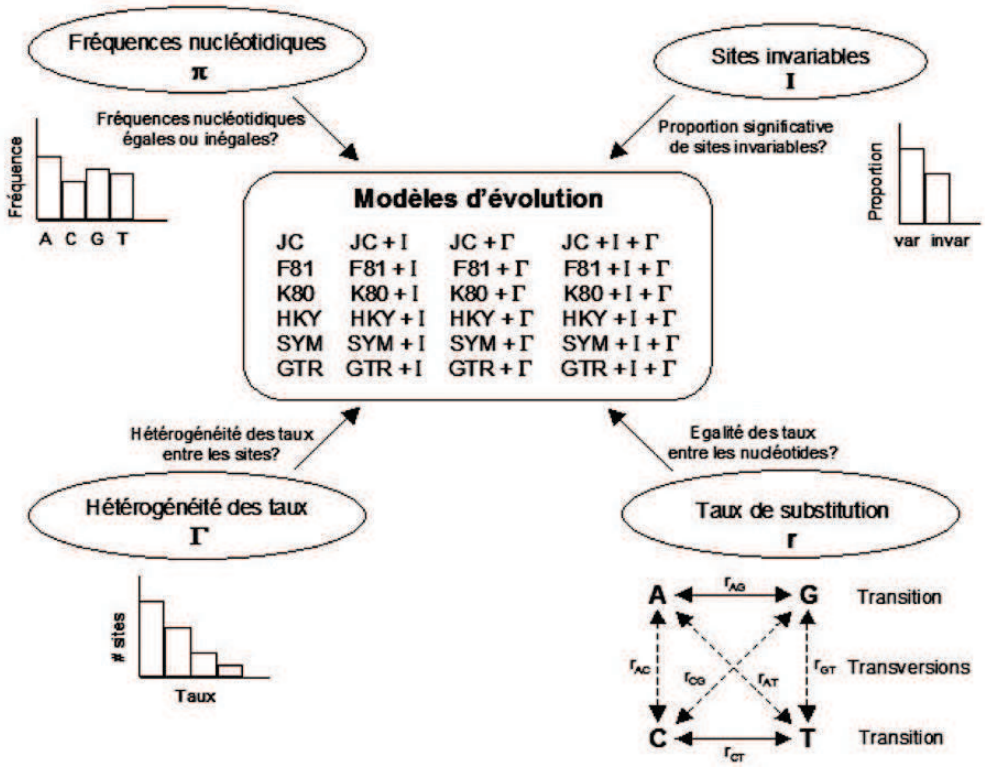


Figure 20 : Composition de quelques modèles d'évolution de séquences nucléotidiques.

Source : Delsuc and Douzery (2004). Avec l'autorisation de Frédéric Delsuc.

Parmi ces méthodes probabilistes, le maximum de vraisemblance cherche à maximiser la vraisemblance. L'arbre sélectionné sera donc celui qui présente la plus forte probabilité d'avoir conduit aux données. L'approche bayésienne diffère du maximum de vraisemblance en utilisant un concept différent des probabilités (Delsuc and Douzery, 2004). Elle donne des probabilités aux paramètres a priori. Elle calcule ensuite les probabilités a posteriori en combinant les probabilités a priori et la vraisemblance. La différence avec le maximum de vraisemblance est qu'il n'y aura pas forcément un seul arbre qui sera sélectionné puisque l'approche bayésienne travaille avec des distributions de probabilités. Une méthode d'échantillonnage des arbres puis la construction d'un consensus majoritaire qui indique le pourcentage de fois où un nœud est majoritairement retrouvé dans l'ensemble des arbres échantillonnés, sont alors réalisées.

Les méthodes de distance présentent l'avantage d'être rapide, car elles n'ont pas à comparer autant d'arbres sous un critère d'optimalité que les méthodes probabilistes par exemple. Elles sont donc utiles pour analyser de larges jeux de données (Yang and Rannala, 2012). De plus les topologies inférées à l'aide de ces méthodes sont en général correctes pour des séquences qui divergent faiblement (Thomas et al., 2010). Cependant, lorsque la divergence entre séquences augmente, ces méthodes sont moins performantes. Elles sont également sensibles aux gaps (Yang and Rannala, 2012). Leur utilité réside également dans le fait que beaucoup de méthodes probabilistes démarrent leur processus à partir d'un arbre de départ et l'améliorent. Cet arbre est souvent généré à partir d'une méthode de distance (Thomas et al., 2010). L'un des principaux inconvénients du maximum de parcimonie est l'impossibilité d'y incorporer un modèle d'évolution de séquences. Il n'est pas possible par exemple de prendre en compte les substitutions multiples à un même site. Cette méthode est particulièrement

BY-NC-ND

sensible à l'attraction des longues branches qui attirent les taxons présentant de longues branches l'un vers l'autre, sans que cela reflète leur véritable relation de parenté. Un autre phénomène qui a tendance à regrouper des taxons alors que ce ne sont pas les plus apparentés, est le biais de composition en bases. La composition en base n'est pas forcément équilibrée et deux taxons qui présentent le même déséquilibre risquent d'être groupés ensemble alors qu'ils ne sont pas apparentés. L'avantage des méthodes probabilistes est leur utilisation de modèles d'évolution explicites dont les paramètres peuvent être évalués. Le temps de calcul de ces méthodes peut cependant être important selon le jeu de données et le modèle utilisé (Yang and Rannala, 2012).

II.3.2. Phylogénie de *T. gondii*

II.3.2.1. Choix du groupe externe

Afin de donner un sens évolutif à l'arbre phylogénétique, il faut l'enraciner. On peut réaliser l'enracinement en incluant un groupe externe au groupe étudié. Il ne doit pas être trop éloigné du groupe étudié afin de permettre de détecter les homologies et donc de comparer. Cependant s'il est trop proche, il peut aussi biaiser l'analyse. La racine de l'arbre se place alors sur le lien qui connecte le groupe étudié au groupe externe. Dans cette étude, une souche d'*Hammondia hammondi* a été choisie pour enraciner l'arbre. Il s'agit d'une espèce différente qui n'est donc pas incluse dans le groupe étudié mais c'est le parasite le plus proche de *T. gondii*. Les séquences de la souche H.H.34 d'*H. hammondi* correspondant aux séquences *GRA6*, *GRA7*, *SAG3*, *UPRT1* et *UPRT7* de *T. gondii* ont été trouvées à partir de GenBank et de ToxoDB. Elles ont été alignées avec le jeu de données en utilisant MUSCLE du logiciel MEGA v5.2.2.

II.3.2.2. Analyses exploratoires : méthodes de distance

Ces analyses ont été effectuées sur l'ensemble du jeu de données. Les haplotypes ont été extraits à l'aide des packages *pegas* et *ape* du logiciel R. Les arbres phylogénétiques ont ensuite été reconstruits avec 3 méthodes de distance : le NJ, le BioNJ et FastMe. Les distances ont été calculées avec le modèle Tamura-Nei (1993) qui prend en compte deux taux différents pour les deux types de transitions (A↔G et C↔T), plus un taux pour les transversions. Mille bootstraps ont été réalisés.

II.3.2.3. Phylogénie : méthode de Maximum de vraisemblance

Les analyses précédentes (II.2.3.1 et II.2.4.1) sur la diversité au sein des marqueurs a mis en évidence que les différentes régions étudiées (gène de ménage/antigènes et régions codantes/régions non-codantes) ne présentaient pas les mêmes caractéristiques. Afin de reconstruire la phylogénie le jeu de données a donc été divisé en 4 : les régions codantes vs non-codantes d'un côté, *GRA6*, *GRA7*, *SAG3* vs *UPRT1* et *UPRT7* de l'autre.

Les analyses phylogénétiques en maximum de vraisemblance ont été réalisées avec le package *phangorn* de R. Pour chaque arbre généré avec une méthode de distance (NJ, BioNj et FastME), les fréquences des bases, les taux de toutes les substitutions possibles, la distribution Gamma, la proportion de sites invariables et la topologie de l'arbre ont été optimisés (m1). A partir de chacun de ces modèles, trois autres (m2, m3 et m4) ont été générés en optimisant différents paramètres pour chaque partition. Le modèle m2 optimisait la topologie et la longueur de branches. Le modèle m3 optimisait le taux de substitutions, la forme

BY-NC-ND

de la distribution Gamma et la proportion de sites invariables. Le modèle m4 optimisait la forme de la distribution Gamma, la proportion de sites invariables, la topologie et la longueur de branches. Les différents modèles ont ensuite été comparés sur la base de leur AIC, le modèle présentant le plus faible AIC étant considéré comme le meilleur. Pour chaque partition de données un arbre phylogénétique a été construit en utilisant le modèle choisi.

II.3.2.4. Consensus des arbres

A partir des quatre arbres précédents générés avec la méthode de maximum de vraisemblance, un arbre consensus a été construit à l'aide du package *ape* du logiciel R (Felsenstein, 2004).

II.3.3. Résultats

II.3.3.1. Les arbres de distances

Les résultats des analyses exploratoires sont présentés dans la figure 21 pour l'arbre construit en NJ, dans la figure 22 pour l'arbre construit en BioNJ et dans la figure 23 pour l'arbre construit en FastME.

Les arbres construits avec ces 3 méthodes ne sont pas identiques entre eux ce qui ne permet pas de conclure avec ces méthodes. De plus, dans les 3 phylogénies, peu d'embranchements sont vraiment bien soutenus, c'est-à-dire avec des bootstraps > 90%. Cependant, il est possible de dégager quelques groupes bien soutenus et de les comparer entre les arbres.

Dans les 3 phylogénies, à partir du groupe externe, deux grands groupes divergent. Les arbres construits avec NJ et BioNJ sont en accord sur cette division : un groupe composé de toutes les souches guyanaises à l'exception de GUY03, de quelques souches du Brésil, de quelques souches du Gabon (GAB07-GAB10, GAB12-14) et d'une souche de Turquie (TUR11) diverge d'un deuxième groupe rassemblant toutes les autres souches. Cette division n'est pas soutenue dans l'arbre construit avec FastME. Cet arbre supporte plutôt une divergence entre un premier groupe composé exclusivement de souches chinoises (CHI04-CHI14) et un deuxième groupe composé de toutes les autres souches. Dans ce dernier groupe, un sous-ensemble diverge composé majoritairement de souches guyanaises. Cependant ce nœud n'est pas vraiment bien soutenu (<80%). Le groupe composé de souches chinoises se retrouve bien supporté dans les deux autres phylogénies bien qu'il ne soit pas à la même place. Un autre regroupement est relativement bien supporté dans l'ensemble des phylogénies (cercle rouge sur les figures 21, 22 et 23) : il se compose de l'haplotype C qui est commun à des souches d'origine géographiques variées (1 des Caraïbes, 1 d'Asie, 2 d'Ethiopie, 4 du Gabon, 2 de Pennsylvanie, 1 e France et 6 du Portugal) et des souches BRN03, GAB05, GAB06 et MIN02. Un autre groupe (cercle bleu sur les figures 21, 22 et 23), composé de CAR04, GUY03 et de la plupart des souches d'Europe, d'Amérique du Nord, d'Ethiopie et de Turquie, est bien supporté dans les 3 phylogénies mais avec quelques incertitudes sur la position relatives de certaines souches en particulier GUY03 et CAR04. Dans ce groupe, l'insertion GAT dans *GRA6* pour les souches GUY03, MIN04-MIN12 et PEN07-PEN09 est aussi présente dans *H. hammondi*.

BY-NC-ND

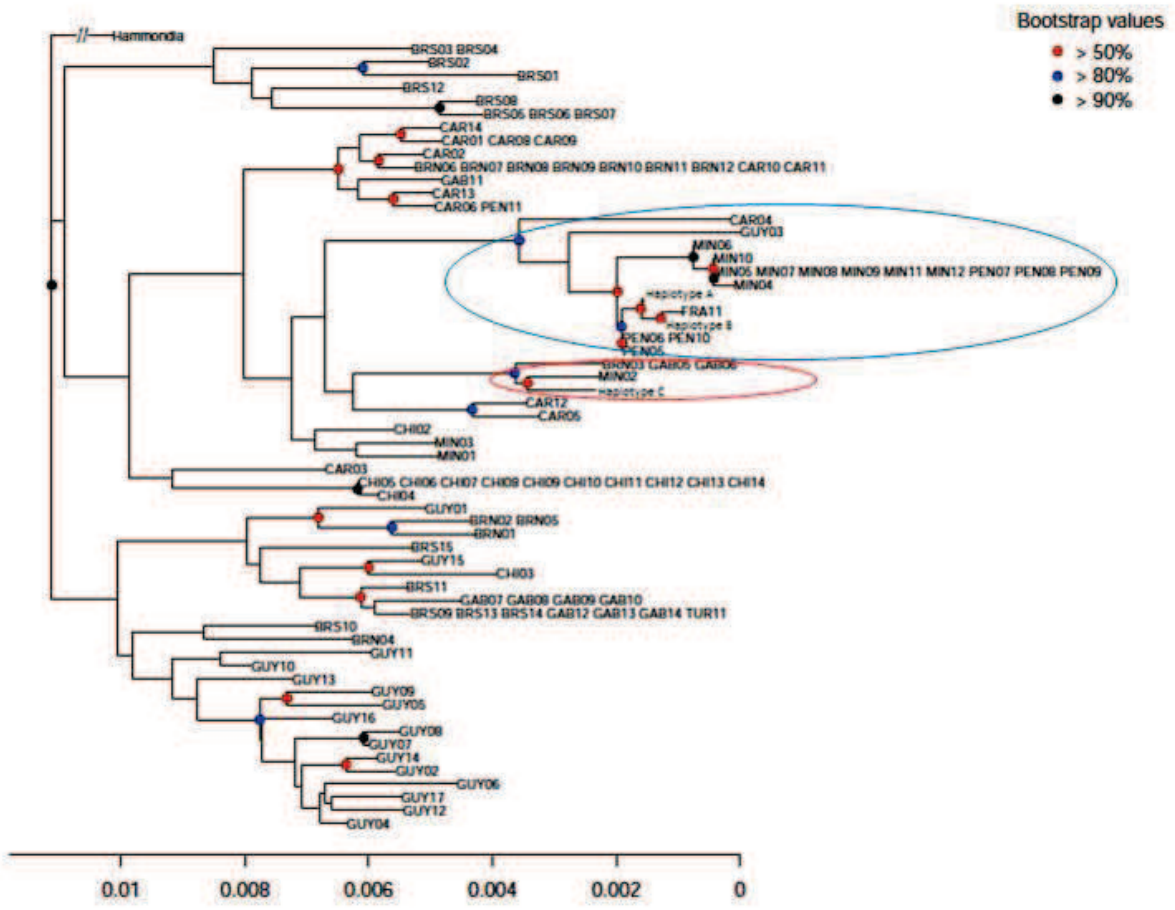


Figure 21 : Arbre NJ enraciné avec *Hammondia hammondi*. Pour la visibilité de la figure, les séquences identiques des haplotypes A, B et C n'ont pas été représentées. Haplotype A : AUS01-07, AUS09, AUS10, ETH03-08, ETH10-13, FRA08-10, POR08, POR09, POR11, TUR02, TUR04, TUR07, TUR08, TUR10 ; Haplotype B : AUS08, ETH09, FRA02-07, FRA12, PEN02-04, POR07, POR10, POR12, TUR01, TUR03, TUR05, TUR06, TUR09, Haplotype C : CAR07, CHI01, ETH01, ETH02, FRA01, GAB01-04, PEN01, PEN12, POR01-06. Les ellipses bleues et rouges représentent deux groupes relativement bien supportés dans l'ensemble des phylogénies.

BY-NC-ND

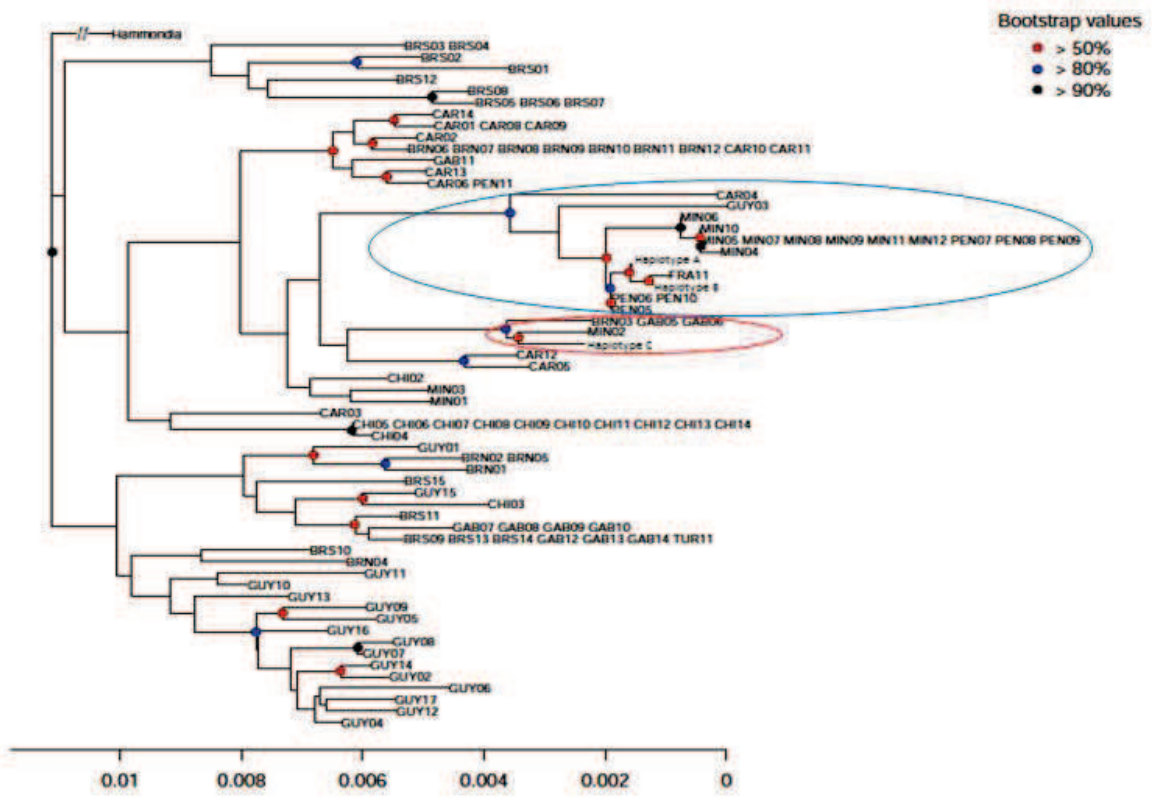


Figure 22 : Arbre BioNJ enraciné avec *Hammondia hammondi*. Pour la visibilité de la figure, les séquences identiques des haplotypes A, B et C n'ont pas été représentées. Haplotype A : AUS01-07, AUS09, AUS10, ETH03-08, ETH10-13, FRA08-10, POR08, POR09, POR11, TUR02, TUR04, TUR07, TUR08, TUR10 ; Haplotype B : AUS08, ETH09, FRA02-07, FRA12, PEN02-04, POR07, POR10, POR12, TUR01, TUR03, TUR05, TUR06, TUR09, Haplotype C : CAR07, CHI01, ETH01, ETH02, FRA01, GAB01-04, PEN01, PEN12, POR01-06. Les ellipses bleues et rouges représentent deux groupes relativement bien supportés dans l'ensemble des phylogénies.

BY-NC-ND

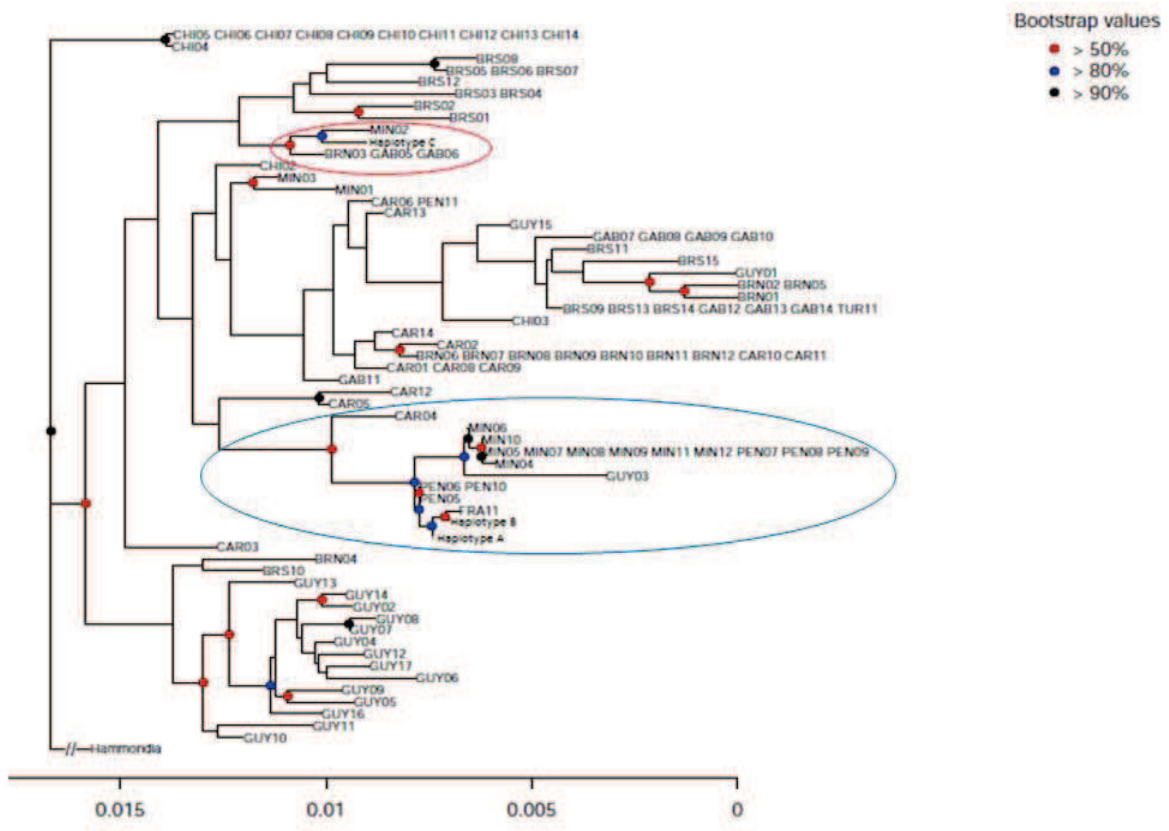


Figure 23 : Arbre en FastMe enraciné avec *Hammondia hammondi*. Pour la visibilité de la figure, les séquences identiques des haplotypes A, B et C n'ont pas été représentées. Haplotype A : AUS01-07, AUS09, AUS10, ETH03-08, ETH10-13, FRA08-10, POR08, POR09, POR11, TUR02, TUR04, TUR07, TUR08, TUR10 ; Haplotype B : AUS08, ETH09, FRA02-07, FRA12, PEN02-04, POR07, POR10, POR12, TUR01, TUR03, TUR05, TUR06, TUR09, Haplotype C : CAR07, CHI01, ETH01, ETH02, FRA01, GAB01-04, PEN01, PEN12, POR01-06. Les ellipses bleues et rouges représentent deux groupes relativement bien supportés dans l'ensemble des phylogénies.

BY-NC-ND

II.3.3.2. Les arbres en maximum de vraisemblance

Les valeurs des AIC pour les différents modèles sont :

Arbre initial : NJ

Modèles	m 1	m 2	m 3	m 4
AIC	17587,2	17393,27	17553,12	17382,38

Arbre initial : BioNJ

Modèles	m 1	m 2	m 3	m 4
AIC	17585,07	17391,1	17551,05	17382,8

Arbre initial : FastME :

Modèles	m 1	m 2	m 3	m 4
AIC	17518,33	17400,69	17483,77	17388

Sur la base du critère d'AIC, le modèle choisi pour construire les arbres en maximum de vraisemblance est le modèle m4 optimisant la forme de la distribution Gamma (Γ), la proportion de sites invariables (I), la topologie et les longueurs des branches. Ce modèle correspond à un modèle GTR + Γ + I, avec des paramètres pouvant varier entre les différentes partitions. L'arbre initial est l'arbre construit en NJ. Les 4 arbres construits à partir des 4 partitions sont présentés dans les figures 24 à 27. Les deux arbres construits sur les régions non-codantes de *GRA6*, *GRA7* et *SAG3* et sur les régions codantes d'*UPRT1* et *UPRT7* sont peu résolus et montrent de larges multifurcations. On peut noter cependant que malgré le peu de résolution de ces arbres, celui sur les régions non-codantes de *GRA6*, *GRA7* et *SAG3* permet de retrouver 2 groupes déjà identifiés dans les arbres de distance : le groupe des souches chinoises ainsi que le groupe constitué de GUY03, CAR04 et de la plupart des souches d'Europe, d'Amérique du Nord, d'Ethiopie et de Turquie. Ces deux groupes sont également bien différenciés par les exons de *GRA6*, *GRA7* et *SAG3*. Sur les régions non-codantes d'*UPRT1* et d'*UPRT7*, le groupe des souches chinoises est aussi bien formé. Cependant, l'autre groupe est mélangé avec d'autres souches et les relations à l'intérieur du groupe sont peu résolues. GUY03 semble être la souche la plus différente dans ce groupe. Les 2 arbres les plus résolus semblent soutenir une position basale des souches d'Amérique du Sud.

BY-NC-ND



Figure 24 : Arbre Maximum de Vraisemblance sur les régions codantes *GRA6, GRA7* et *SAG3*, enraciné avec *Hammondia hammondi*. Haplotype A : AUS01-07, AUS09, AUS10, ETH03-08, ETH10-13, FRA08-10, POR08, POR09, POR11, TUR02, TUR04, TUR07, TUR08, TUR10 ; Haplotype B : AUS08, ETH09, FRA02-07, FRA12, PEN02-04, POR07, POR10, POR12, TUR01, TUR03, TUR05, TUR06, TUR09, Haplotype C : CAR07, CHI01, ETH01, ETH02, FRA01, GAB01-04, PEN01, PEN12, POR01-06

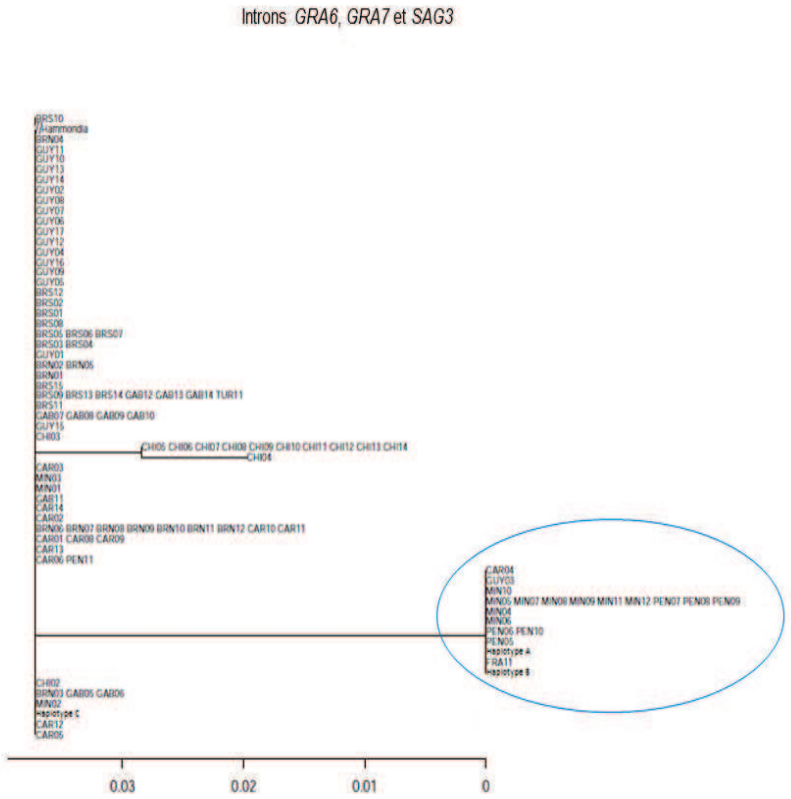


Figure 25 : Arbre Maximum de vraisemblance sur les régions non-codantes de *GRA6, GRA7* et *SAG3*, enraciné avec *Hammondia hammondi*. Haplotype A : AUS01-07, AUS09, AUS10, ETH03-08, ETH10-13, FRA08-10, POR08, POR09, POR11, TUR02, TUR04, TUR07, TUR08, TUR10 ; Haplotype B : AUS08, ETH09, FRA02-07, FRA12, PEN02-04, POR07, POR10, POR12, TUR01, TUR03, TUR05, TUR06, TUR09, Haplotype C : CAR07, CHI01, ETH01, ETH02, FRA01, GAB01-04, PEN01, PEN12, POR01-06

BY-NC-ND

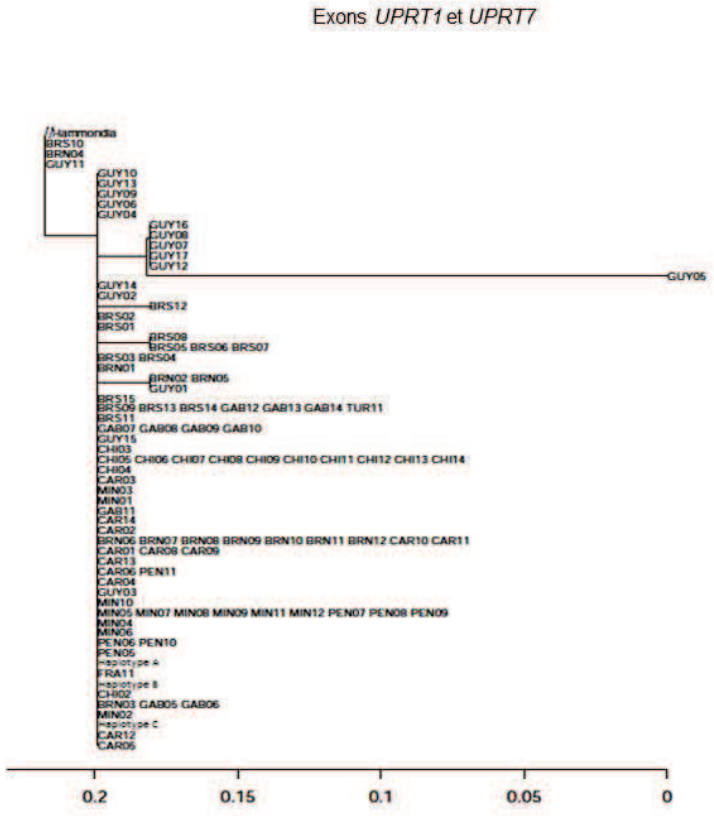


Figure 26 : Arbre Maximum de Vraisemblance sur les régions codantes d'*UPRT1* et *7*, enraciné avec *Hammondia hammondi*. Haplotype A : AUS01-07, AUS09, AUS10, ETH03-08, ETH10-13, FRA08-10, POR08, POR09, POR11, TUR02, TUR04, TUR07, TUR08, TUR10 ; Haplotype B : AUS08, ETH09, FRA02-07, FRA12, PEN02-04, POR07, POR10, POR12, TUR01, TUR03, TUR05, TUR06, TUR09, Haplotype C : CAR07, CHI01, ETH01, ETH02, FRA01, GAB01-04, PEN01, PEN12, POR01-06.

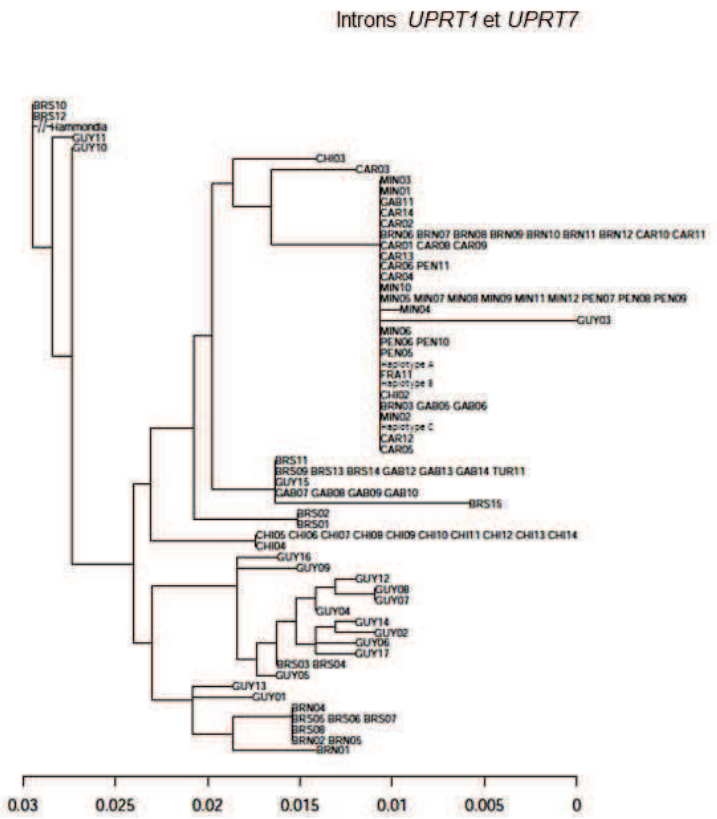


Figure 27 : Arbre Maximum de Vraisemblance sur les régions non-codantes d'*UPRT1* et *7*, enraciné avec *Hammondia hammondi*. Haplotype A : AUS01-07, AUS09, AUS10, ETH03-08, ETH10-13, FRA08-10, POR08, POR09, POR11, TUR02, TUR04, TUR07, TUR08, TUR10 ; Haplotype B : AUS08, ETH09, FRA02-07, FRA12, PEN02-04, POR07, POR10, POR12, TUR01, TUR03, TUR05, TUR06, TUR09, Haplotype C : CAR07, CHI01, ETH01, ETH02, FRA01, GAB01-04, PEN01, PEN12, POR01-06

BY-NC-ND

II.3.3.3. Le consensus

Le consensus des arbres montre une multifurcation à la base de l'arbre. Elle contient 35 des 44 souches d'Amérique du Sud, 12 souches de Chine, et 7 souches d'Afrique (Figure 28).

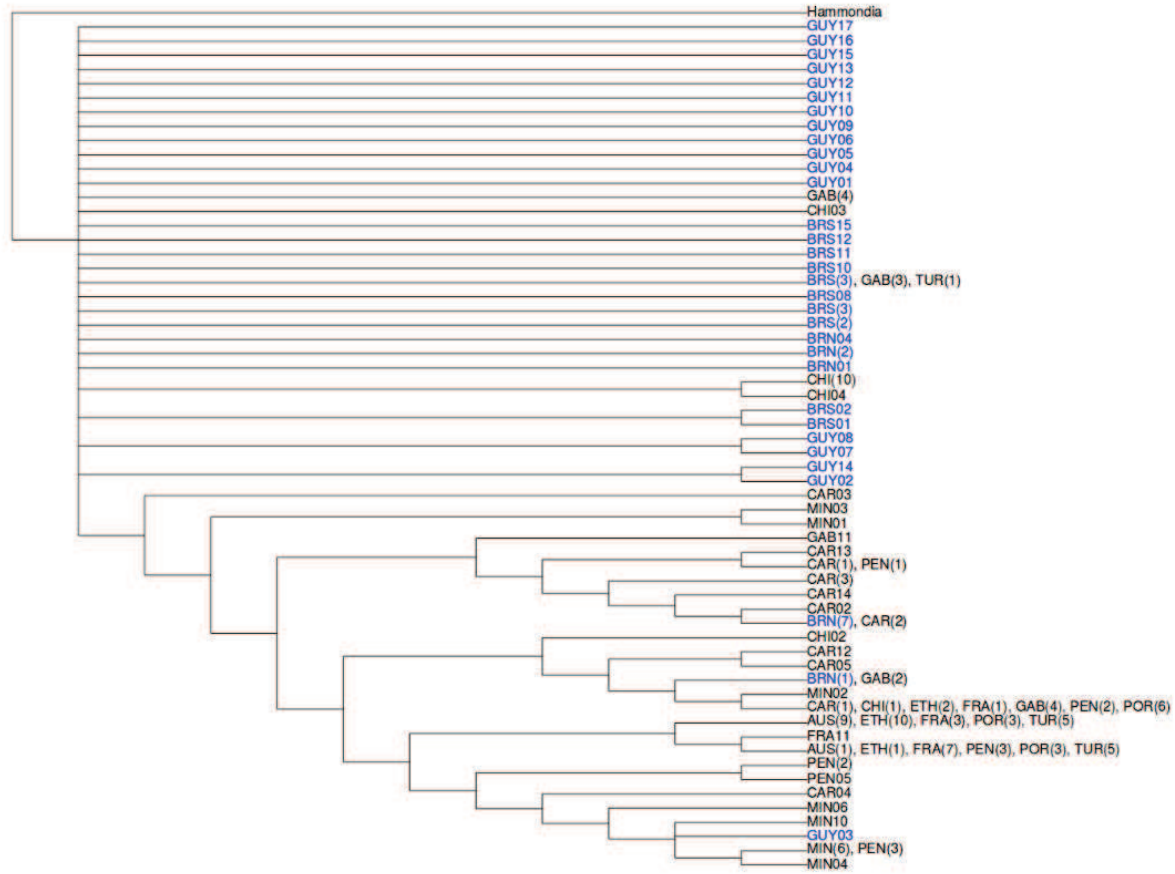


Figure 28 : Consensus des 4 arbres correspondants aux partitions suivantes régions codantes/ régions non-codantes, *GRA6*, *GRA7*, *SAG3* / *UPRT1*, *UPRT7.* Les noms en bleu représentent des souches d'Amérique du Sud.

II.3.4. Discussion

Les analyses phylogénétiques ne permettent pas de décrire avec précision les relations de parenté entre les différentes souches de *T. gondii*. En effet, les différents arbres reconstruits ne sont pas cohérents entre eux, que ce soit les arbres construits à partir du jeu de données complet mais avec des méthodes différentes (NJ, BioNJ, FastME et ML) ou les arbres construits avec la même méthode mais sur des jeux de données différents. Cependant, la consistance de certains groupes dans les différents arbres bien que leur localisation précise ne soit pas déterminée dans la phylogénie, permet de noter quelques tendances. Ainsi la plupart des souches de Guyane semblent se regrouper entre elles, ce qui est aussi le cas pour la plupart des souches chinoises et européennes. Les phylogénies semblent confirmer que peu d'échanges se produisent entre l'Amérique du Sud et l'Amérique du Nord. Cependant, le rapprochement d'une souche de Guyane, isolée à partir d'un jaguar (GUY03) avec des souches d'Amérique du Nord suggèrent que bien que rares, des échanges semblent avoir lieu. La présence de souches Africaines dans des groupes composés en particulier de souches

BY-NC-ND

sud-américaines semble aussi confirmer des échanges entre ces deux continents. Cependant, il s'agit exclusivement de souches du Gabon, les souches de l'Ethiopie n'ayant aucun haplotype commun avec les souches sud-américaines. Les échanges entre l'Amérique du Sud et l'Afrique concerneraient ainsi surtout la partie Ouest de l'Afrique. Il existe quelques souches du Gabon qui semblent être plus proches des souches d'Europe et d'Ethiopie. La partie Est de l'Afrique serait elle, plus susceptible d'échanger avec l'Eurasie. La Turquie semble également présenter une certaines diversité dans ces échanges avec les autres zones. En effet, toutes les souches à l'exception de TUR11 ont été isolées aux alentours de Izmir, à l'extrémité Est de la Turquie et se regroupent entre elles. TUR11 fait exception car elle se retrouve dans le groupe comprenant des souches d'Amérique du Sud et du Gabon. Cependant, d'autres analyses, comportant des souches supplémentaires de Turquie et plus particulièrement de la région d'Ankara seraient nécessaires pour vérifier s'il y a une certaine structure génétique en Turquie.

Le manque de consistance entre les différentes phylogénies montre qu'il est difficile de reconstruire des phylogénies intra-espèces. Pour déterminer les raisons de ce manque de consistance, il pourrait être intéressant d'évaluer la discordance entre les différentes topologies. La présence de taxons dont la position phylogénétique n'est pas connue de façon certaine pourrait être à l'origine de la faible résolution du consensus. Il serait donc interessant de déterminer quels sont ces taxons et de les éliminer de l'analyse afin d'avoir des arbres plus robustes et donc de faire des estimations plus précises du TMRCA et du taux de mutations. Enfin, au lieu de construire un arbre consensus à partir de plusieurs arbres, une approche alternative pourrait consister à utiliser un alignement de base résultant de la concaténation des alignements des différents loci. Les gènes choisis pour l'analyse ont été choisis selon un critère de polymorphisme, afin de pouvoir déterminer des diversités génétiques intra-populationnelles malgré la tendance importante qu'a *T. gondii* de se propager de façon clonale. Cependant il est recommandé en phylogénie d'utiliser des séquences d'ADN neutres du point de vue de la sélection, ce qui n'est pas forcément le cas ici, en particulier avec *GRA6*, *GRA7* et *SAG3* qui contiennent majoritairement des régions codantes. De plus les recombinaisons peuvent aussi brouiller le signal phylogénétique. Le choix des gènes semble être une limite pour ces analyses phylogénétiques. Il serait intéressant, de les réaliser avec d'autres régions d'ADN, plus neutres, tels que des introns. Ce travail en comporte 2, *UPRT1* et *UPRT7*. Cependant, ils font partie du même gène, autre limite de ce type d'analyse. Toutes les régions d'un génome n'évoluent pas de la même façon et il est donc nécessaire de prendre des régions qui se trouvent à différents endroits afin d'avoir plusieurs régions représentées. Afin de palier au problème des recombinaisons, il serait aussi intéressant de réaliser les mêmes analyses avec des régions d'ADN ne provenant pas du génome nucléaire mais d'organites telles que les mitochondries ou l'apicoplaste. En effet, ces organites présents dans *T. gondii* sont hérités de façon uni-parentale. Leurs génomes sont donc moins sensibles aux recombinaisons et pourraient donc être plus informatifs d'un point de vue phylogénétique.

Les analyses précédentes qui montrent une diversité génétique plus élevées en Amérique du Sud et qui suggèrent qu'une majorité des souches d'Amérique du Sud se situeraient en position basale dans la phylogénie, tendent à montrer que l'Amérique du Sud serait le lieu d'origine des souches actuelles de *T. gondii*. Cependant, la démonstration d'une diversité plus élevée dans un lieu étant seulement un indice pouvant suggérer qu'il s'agit du lieu d'origine et la phylogénie n'étant pas assez robuste, une analyse spatiale a été réalisée.

BY-NC-ND

II.4. Analyse spatiale

II.4.1. Principe de la relation diversité génétique/distance géographique

La diversité génétique peut être influencée par des fluctuations de la taille des populations. Ces fluctuations peuvent avoir lieu au cours de goulots d'étranglements, évènements qui se produisent lorsque quelques individus et donc quelques génotypes présents dans la population survivent, ou au cours d'effet de fondation, processus au cours duquel, seuls quelques individus migrent à partir d'une population de départ pour en fonder une nouvelle. Dans les deux cas, seuls quelques génotypes subsistent dans la nouvelle population, conduisant à une perte de la diversité génétique. Au cours de l'expansion d'un organisme, plusieurs effets de fondation vont donc avoir lieu, quelques migrants quittant à chaque fois les populations nouvellement formées pour en créer de nouvelles, conduisant à la répartition de l'organisme et à une perte successive de la diversité génétique (Figure 29).

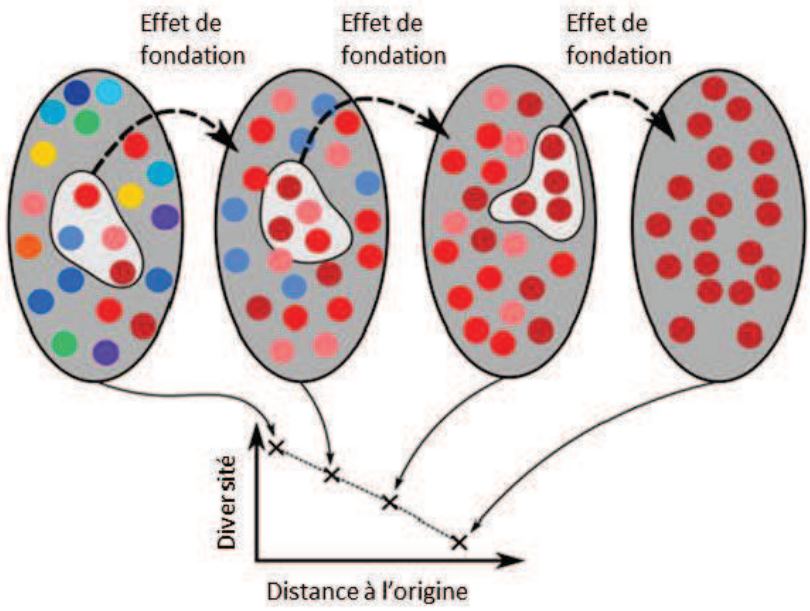


Figure 29 : Les migrations successives conduisent à une perte progressive de la diversité de la population. Les ellipses représentent des populations et les ronds des organismes composant ces populations. Les couleurs représentent la diversité présente dans la population.

Source : Adapté de Mita and Jombart (2015)

II.4.2. Détermination de l'origine géographique de *Plasmodium falciparum*

Ce principe d'une diminution de la diversité génétique lorsque les distances à l'origine augmentent a été utilisé lors d'une étude pour déterminer l'origine de *Plasmodium falciparum.* Il s'agit d'un parasite, faisant partie du groupe des Hémosporidies des Apicomplexes (Figure 2). Il se distribue dans la zone tropicale et est transmis à l'homme par la piqure de certaines espèces d'anophèles, causant le paludisme. *P. falciparum* se reproduit de façon sexuée chez l'anophèle qui est donc l'hôte définitif. 519 échantillons, provenant de patients infectés, se

BY-NC-ND

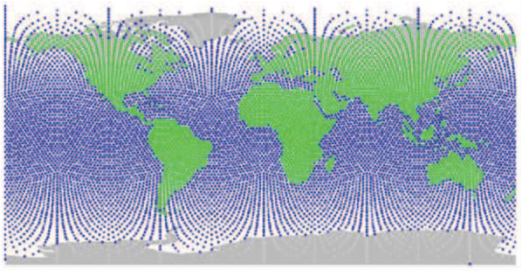
répartissant dans 9 populations couvrant la zone de distribution *de P. falciparum* (à l'exception des Amériques*)* ont été sélectionnés. Pour chaque isolat 9 gènes de ménage ont été séquencés c'est à dire un total 5043 bp pour 63 SNPs. A partir de ces gènes, la diversité génétique a été calculée dans chaque population. La corrélation entre les diversités génétiques des différentes populations et les distances géographiques à l'origine ont été calculées pour 312 origines hypothétiques. L'Afrique sub-saharienne a été l'origine du pathogène pour laquelle la meilleure corrélation négative a été trouvée. La structure de la diversité génétique de *P. falciparum* est semblable à celle de la population humaine. Un modèle génétique a donc été réalisé afin de dater l'expansion du parasite à partir de l'Afrique. L'ensemble des résultats suggère que les humains étaient déjà infectés avant leur sortie d'Afrique et que le parasite s'est répandu avec eux à travers les tropiques, à l'exception de l'Amérique du Sud qui a été colonisé beaucoup plus tard par *P. falciparum* (Tanabe et al., 2010).

II.4.3. Application au cas de *Toxoplasma gondii*

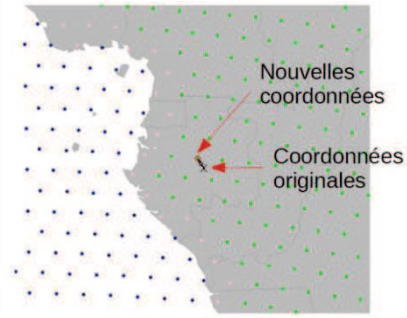
L'objectif de cette partie était d'appliquer le modèle précédent au cas de *T. gondii*. Ce parasite se distingue des autres apicomplexes par un spectre d'hôte et une distribution géographique très larges. A la différence de *P. falciparum*, les tropiques ne sont pas la seule zone à prendre en compte et plusieurs hôtes intermédiaires ont pu influencer la répartition du parasite en plus des félidés. La première étape a été de choisir l'indice de diversité génétique le plus adapté pour ce parasite. Dans le cas de *T. gondii* qui présente une forte part de reproduction clonale, une diversité élevée essentiellement due à des indices qui favorisent un faible nombre d'haplotypes très différents nous a semblé artificielle. La richesse haplotypique est donc l'indice qui est apparu comme le plus approprié pour estimer la diversité génétique de *T. gondii*. L'analyse spatiale a été effectuée avec le package *geoGraph* (http://thibautjombart/geograph) du logiciel R qui fournit la méthode utilisée dans l'étude de *P. falciparum*. Une grille comportant approximativement 40 000 points uniformément répartis sur la carte a été utilisée (Figure 30.a). Chaque point possède un « habitat » afin de distinguer les points se trouvant sur les terres de ceux se trouvant dans les océans et les mers. Afin de placer les populations sur la carte, les coordonnées géographiques de chaque population ont été calculées comme étant la moyenne des coordonnées des individus la composant. Chaque population a ensuite été associée au point « terrestre » le plus proche de ses coordonnées (Figure 30.b). Les migrations des oiseaux ont été considérées comme trop rares dans la direction Est-Ouest pour expliquer une expansion de *T. gondii* par ce moyen à travers les océans et le transport par bateau trop récent. Ainsi, le modèle utilisé considère que seul le passage par les terres était possible. Par conséquent, seuls les points « terrestres » ont été reliés entre eux (Figure 30.c). Les liens entre les points symbolisent les routes que le parasite a pu emprunter. Afin de définir des origines géographiques à tester, 1800 combinaisons de longitudes et de latitudes régulièrement espacées ont été utilisées afin de générer 433 points non-redondants uniformément répartis sur les parties terrestres de la grille utilisée par geoGraph (Figure 30.d). La figure 30 illustre certaines des étapes de cette analyse. Pour chaque origine, les chemins les plus courts vers chaque population ont été calculés avec l'algorithme dijkstra du package de RBGL de R. Les coûts étaient uniformes le long des branches c'est-à-dire qu'aucun chemin n'a été considéré comme plus « probable ». Les distances correspondantes ont été calculées en kilomètres. Les corrélations entre les diversités génétiques et ces distances à l'origine ont été calculées pour chaque origine testée. La corrélation la plus négative a ensuite été

BY-NC-ND

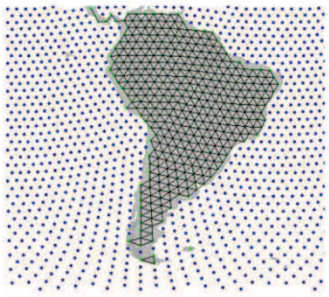
recherchée et a été considérée comme correspondant à l'origine géographique la plus probable.



A: Utilisation d'une grille comportant 40 000 points uniformément répartis. Les points "terrestres" (en vert) sont différenciés des points dans les océans et les mers (en bleu).



B: Après calcul de leurs coordonnées géographiques, les populations sont reliées au point "terrestre" le plus proche.



C: Seuls les points terrestres ont été reliés entre eux. Le modèle empêche donc le passage par les mers et les océans.



D: 433 points uniformément répartis sur les terres ont été générés. Ils représentent les origines hypothétiques testées.

Figure 30 : Représentations des premières étapes de l'analyse spatiale.

BY-NC-ND

II.4.4. Résultats

L'Amérique du Sud et plus précisément la Colombie serait l'origine la plus probable (r=0,81, p=0,9x10^{-4}). De plus, cette analyse permet de reconstruire les routes ancestrales de migration de *T. gondii*. Les résultats suggèrent que ce pathogène se serait répandu en Amérique et aurait ensuite colonisé l'Asie et l'Europe en traversant le détroit de Béring. Il serait ensuite entré en Afrique en suivant deux voies de migration différentes (Figure 31).

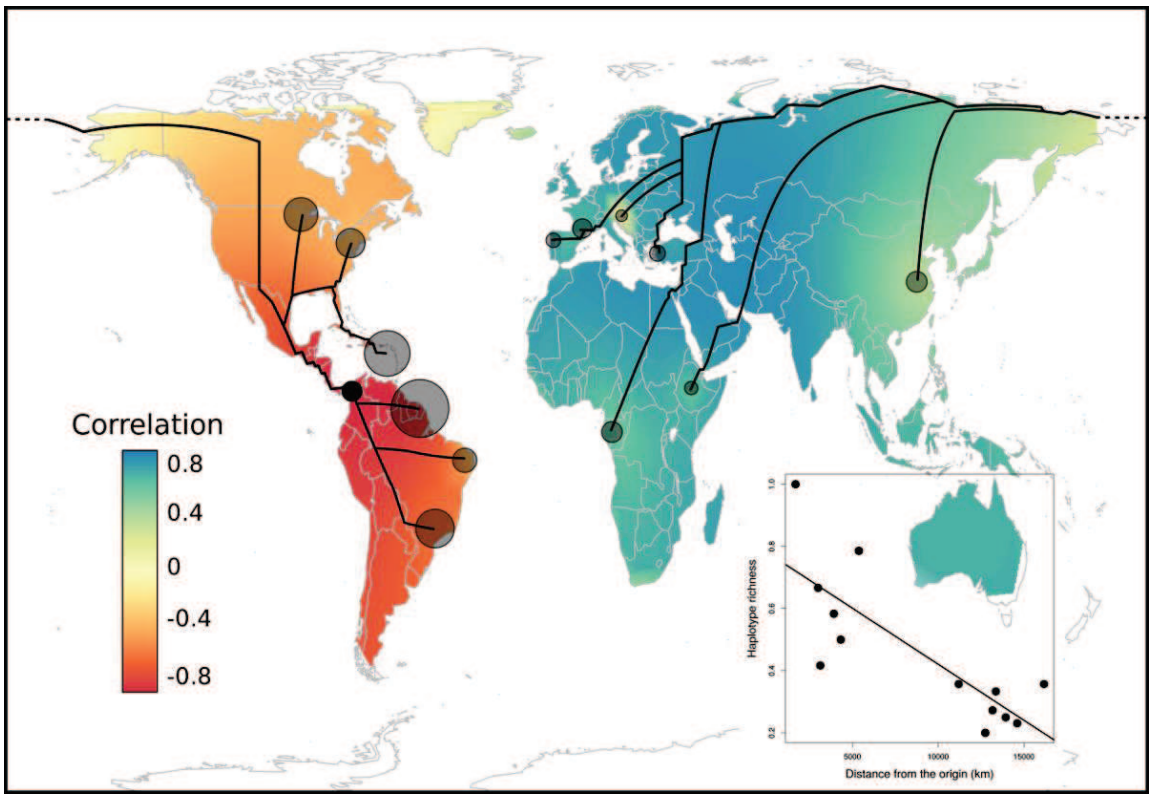


Figure 31 : Répartition historique de *Toxoplasma gondii* inférée à partir des 5 marqueurs. La taille des ronds noirs transparents est proportionnelle à la valeur de la diversité en haplotypes. La diversité en haplotypes des différentes populations a été corrélée aux distances géographiques à partir d'origines variées. Le dégradé de couleur représente ces corrélations. Les régions en rouge foncé représentent les zones comprenant des origines géographiques testées pour lesquelles lorsqu'on s'en éloigne, les diversités génétiques des populations diminuent fortement. L'origine la plus vraisemblable (rond plein noir) correspond à celle pour laquelle il y a la plus forte corrélation négative entre les diversités génétiques et les distances à cette origine.
Les chemins les plus courts à partir de l'origine vers les diverses populations sont indiqués en lignes pleines noires.

BY-NC-ND

II.4.5. Discussion

Ce résultat confirme que l'Amérique du Sud serait l'origine géographique de *T. gondii*, ce qui était suggéré par la tendance des souches sud-américaines d'être en position basale lors des analyses phylogénétiques. Il supporte aussi l'étude réalisée à partir d'isolats de poulets (Lehmann et al., 2006). En plus de se baser sur une diversité plus importante en Amérique du Sud pour en déduire qu'il s'agit sans doute de l'origine, cette analyse met en évidence qu'il y a une corrélation significative entre les distances des différentes populations à l'origine et les diversités génétiques, confortant ainsi l'hypothèse de départ. Elle permet en outre de déduire les voies ancestrales de dissémination du parasite. Le modèle utilisé montre qu'il y a eu vraisemblablement deux introductions différentes en Afrique, l'une passant par le sud de l'Arabie Saoudite et le Yémen jusqu'à la population d'Ethiopie et l'autre passant plus vers le Nord par la Turquie, la Syrie puis le nord de l'Afrique et ensuite l'Ouest. Ces différentes voies ancestrales, associées aux plus récents échanges de souches entre l'Amérique du Sud et l'Afrique de l'Ouest via le transport maritime, pourrait expliquer la différenciation génétique existante entre l'Est et l'Ouest de l'Afrique. Cependant, une limite du modèle présentant la richesse haplotypique en fonction de la distance à l'origine est que les points ne sont pas répartis de façon homogène mais forment deux groupes bien séparés. Ces groupes correspondent aux populations d'Amérique d'une part et aux populations d'Eurasie et d'Afrique d'autre part, deux groupes de populations séparés par le détroit de Béring.

II.5. Datations

II.5.1. Objectif et méthodes de datation

Dans le cadre de ce travail, l'objectif de cette partie était de déterminer l'âge du plus récent ancêtre commun (TMRCA) des lignées actuelles de *T. gondii*. Le TMRCA est le temps auquel les allèles coalescent vers un ancêtre unique (Pettengill, 2015). Deux approches différentes ont été utilisées : une datation moléculaire et la méthode de coalescence.

BY-NC-ND

II.5.1.1. Datation moléculaire

Les datations moléculaires reposent sur le principe selon lequel les changements dans les macromolécules biologiques se produisent de façon régulière. C'est en 1965 qu'Emile Zuckerkandl et Linus Pauling pensent à utiliser ce concept pour déterminer des âges de divergence. Lorsqu'un point de calibration est connu, c'est-à-dire une date de divergence entre deux séquences, il est possible à partir du degré de divergence entre les deux séquences, d'en déduire le taux de substitutions. Dans le cas d'une horloge moléculaire parfaite, c'est-à-dire dans le cas où le taux d'évolution est constant, il est alors possible de le réutiliser pour calculer la date de divergence aux différents nœuds de l'arbre, en utilisant les degrés de divergence (Fig. 32). Cette approche nécessite cependant d'avoir une phylogénie assez robuste et elle est souvent utilisée pour calculer les dates de divergence entre espèces, en utilisant des points de calibration paléontologiques (Douzery et al., 2006).

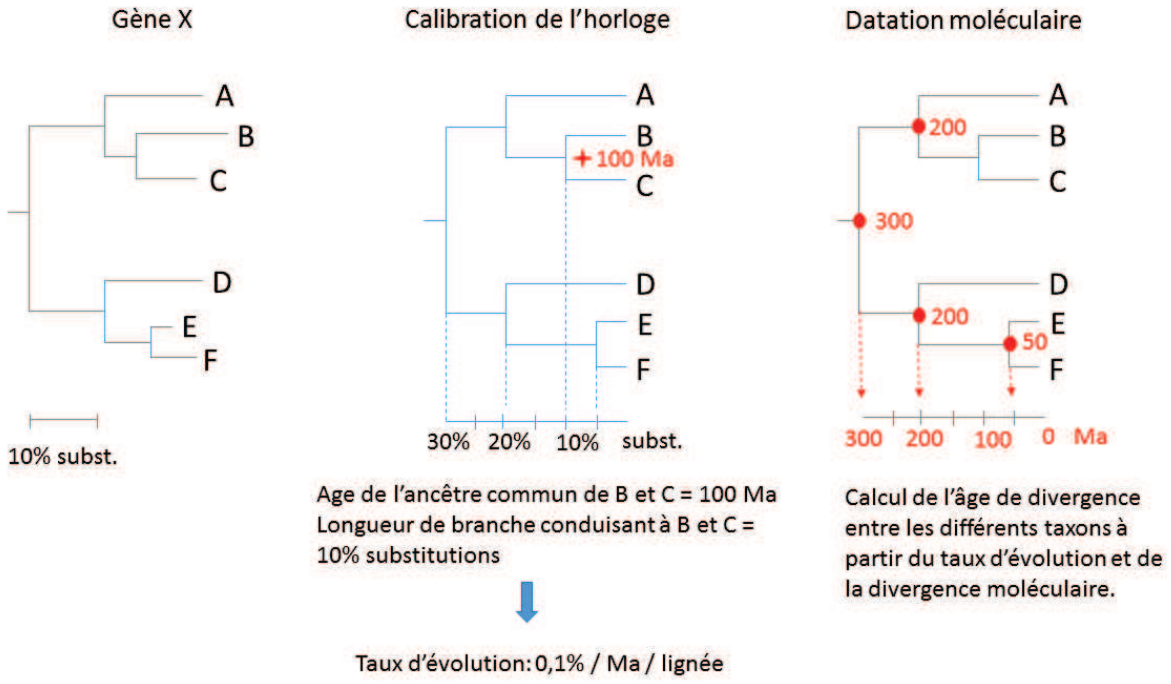


Figure 32 : Principe de la datation moléculaire : Un gène X est considéré pour plusieurs taxons (A-F). L'arbre de gauche peut être contraint à suivre une horloge moléculaire. Tous les taxons sont alors à la même distance de la racine (arbre du milieu). Les échelles grise et bleue en bas des arbres sont formulées en pourcentage de substitutions nucléotidiques. Chacune des branches conduisant à B et C représentent 10% de substitutions. Ainsi le taux de substitution peut être calculé et est égal à 0,1%/Ma/lignée. En utilisant ce taux, les dates de divergences des autres taxons peuvent être calculées (arbre de droite). Par exemple E et F présentent 5% de divergence moléculaire. En considérant que le taux de substitutions est constant et vaut 0,1%/Ma/lignée, la date de divergence entre E et F est de 5/0,1=50 Ma.

Source : Adapté de Douzery et al.(2006)

Cette approche est dépendante de plusieurs facteurs. Les calculs de datation dépendent des valeurs des points de calibration. Ainsi une erreur dans l'étape de calibration de l'horloge moléculaire peut fausser plus ou moins grandement les estimations des dates de divergence.

BY-NC-ND

Les gènes utilisés influencent également ce type de datation. Ils n'évoluent en effet pas tous à la même vitesse et selon les gènes choisis, la dates peuvent être différentes. Enfin, cette approche suppose que les différences de longueurs de branches dans les arbres ne sont pas statistiquement significatives et que donc les arbres peuvent être contraints à suivre une horloge moléculaire. Il a été montré que c'est rarement le cas dans la nature, les taux de substitutions sont rarement constants. Contraindre un arbre à suivre l'horloge moléculaire peut donc entraîner une distorsion trop importante et entraîner des estimations des âges fausses. (Douzery et al., 2006).

II.5.1.2. Méthode de coalescence

La coalescence est définie comme la fusion des lignées ancestrales en remontant le temps. L'idée de base est de partir des lignées échantillonnées et de remonter le temps. Lorsque ces lignées se rejoignent, on dit qu'elles coalescent. Lorsque toutes les lignées coalescent en une seule, il s'agit de l'ancêtre commun le plus récent (MRCA) (Figure 33).

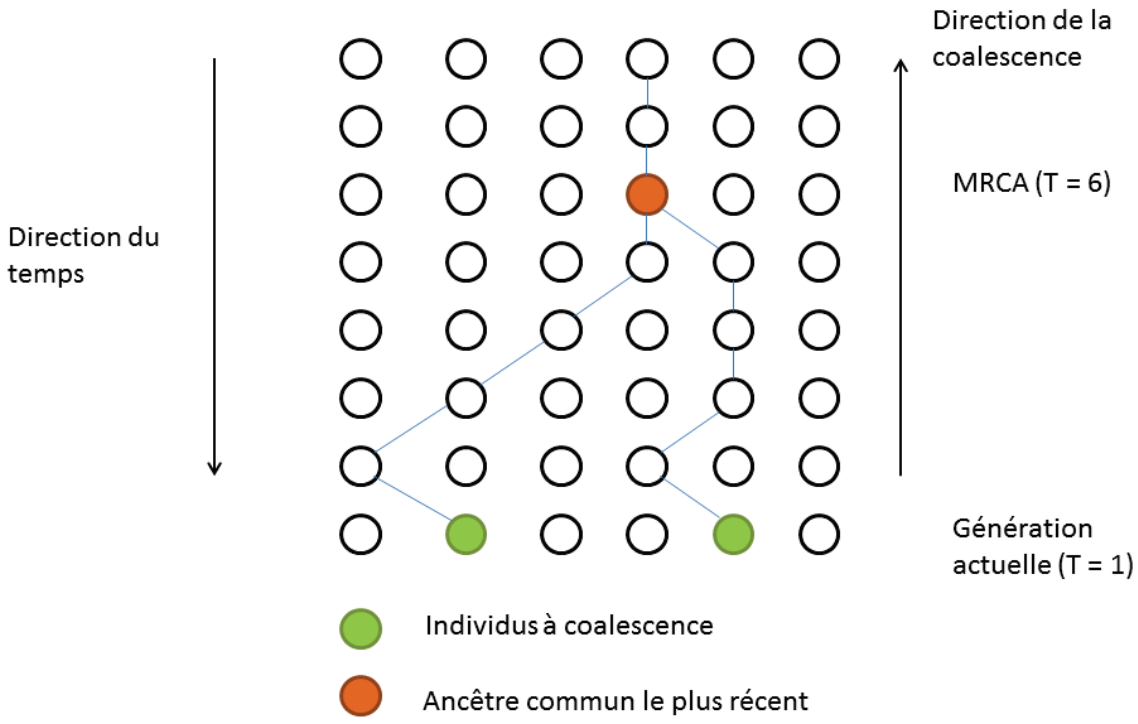


Figure 33 : Processus de coalescence. Il commence avec la génération actuelle (en bas) et remonte dans le temps vers l'ancêtre commun le plus récent (MRCA) (rond orange). Ici 2 individus (ronds verts) coalescent à la sixième génération en remontant dans le temps.

Source : Adapté de Yang et al. (2014)

Le coalescent est un processus stochastique. Il s'agit d'un modèle de génétique des populations qui permet de construire et d'analyser des généalogies au hasard. Contrairement

BY-NC-ND

à la phylogénie, les méthodes de coalescence ne cherchent pas à déterminer un arbre. Au niveau d'une population, il peut être plus pertinent d'utiliser la coalescence pour comprendre son histoire plutôt que la phylogénie. L'histoire d'une population ne suit pas forcément un arbre phylogénétique, notamment parce que la migration peut jouer un rôle important. Les méthodes de coalescence reposent sur le traitement aléatoire des généalogies et l'utilisation des arbres générés afin d'estimer les paramètres du processus généalogique aléatoire qui a créé chaque arbre. L'analyse de la vraisemblance des modèles évolutifs est également possible.

II.5.2. Datation du TMRCA des lignées actuelles de *T. gondii*

Les analyses pour dater le TMRCA ont été réalisées sur les régions non-codantes d'*UPRT*. Les deux types d'approches ont nécessité une estimation du taux de mutation.

II.5.2.1. Calcul du taux de mutation

Afin d'estimer le taux de mutation, les distances entre la séquence d'*H. hammondi* et chaque séquence de *T. gondii* ont été calculées en utilisant le modèle de Tamura-Nei. La moyenne de ces distances a été considérée comme une estimation du nombre de mutations qui ont eu lieu depuis la séparation des 2 espèces. Le taux de mutation a ensuite été estimé selon la formule suivante :

$$\mu = \frac{Nombre\ de\ mutations}{2 \times temps\ de\ divergence}$$

Ici le temps de divergence est celui entre *H. hammondi* et *T. gondii*, soit 11 Ma (Morrison, 2005).

L'écart-type de cette estimation a été calculé en utilisant la variance du ratio de deux variables aléatoires qui sont le nombre de mutations et le temps de divergence. Deux intervalles de confiance ont été calculés de cette façon, l'un en prenant de façon arbitraire une variance de 10^{12} pour le temps de divergence, l'autre en considérant que la date de divergence était connue de façon certaine.

II.5.2.2. Datation moléculaire

Afin de calculer le TMRCA, la plus grande distance entre les haplotypes de *T. gondii* a été prise en compte selon la formule :

$$TMRCA = \frac{0{,}5 \times maxD}{\mu}$$

Où maxD est la distance maximale entre les haplotypes de *T. gondii* et μ le taux de mutation calculé précédemment.

II.5.2.3. Méthode de coalescence

Le TMRCA a également été calculé avec une approche de coalescence. Elle permet d'obtenir une estimation du TMRCA qui est égal à deux fois la taille efficace de population Ne. Elle

BY-NC-ND

nécessite d'estimer le paramètre de population P = 2μNe. Ce paramètre a été estimé avec deux approches. La première a utilisé la méthode de Monte-Carlo par chaînes de Markov à l'aide du package *coalescentMCMC* du logiciel R. La deuxième approche a utilisé la diversité nucléotidique (π) à l'aide du package *pegas*.

II.5.3. Résultats

II.5.3.1. Taux de mutation

Le taux de mutation était de 6,19 x 10^{-9}. L'intervalle de confiance calculé en prenant en compte une variance pour le temps de divergence entre *T. gondii* et *H. hammondi* est de 6,19 x 10^{-9} ± 1,48 x 10^{-9}. En considérant que la date de séparation des deux espèces est connue de façon certaine, c'est-à-dire sans prendre en compte de variance, l'intervalle de confiance est de 6,19 x 10^{-9} ± 1,37 x 10^{-9}.

L'incorporation d'une variance du temps de divergence augmente légèrement l'intervalle de confiance.

II.5.3.2. Les datations du TMRCA

Avec la méthode de la datation moléculaire, le plus récent ancêtre commun a été daté à 1,59 Ma (IC 95% : [0,00–3,46]). Pour les approches de coalescence, celle utilisant la méthode MCMC a estimé que le MRCA daterait de 1,26 Ma (IC 95% [0,94–1,57]) tandis que la méthode utilisant la diversité nucléotidique a estimé le TMRCA à 1,20 Ma (IC 95% [0,00–2,44]).

II.5.4. Discussion

Les 3 estimations du TMRCA réalisées avec deux approches différentes (datation moléculaire et coalescence) sont cohérentes entre elles allant de 1,20 à 1,59 Ma. Cependant, seul un intervalle de confiance semble raisonnable, celui de la méthode MCMC. En effet les deux autres sont très grands et la borne inférieur est à 0. Ce résultat est cohérent avec les calculs faits dans des études antérieures lorsque les taux de mutations utilisés étaient ceux de *P. falciparum* et non pas ceux de *T. gondii* (Khan et al., 2007; Su et al., 2003). Il a cependant été suggéré que les taux de substitutions de ces deux espèces qui se sont séparées il y a au moins plusieurs centaines de millions d'années n'étaient pas forcément les mêmes et il est donc recommandé d'utiliser des taux spécifiques à *T. gondii* (Morrison, 2005). Dans l'étude de Su et al. (2003), le taux de mutation spécifique à *T. gondii* a également été calculé sur la base des introns *ACT1*, *ATUB* et *MIC2* et utilisé dans les datations. Cependant, les résultats étaient différents avec un TMRCA, d'environ 180 000 ans et non plus 1 million d'années. Le taux de mutation trouvé était en moyenne de 2,12 x 10^{-8}, une valeur plus élevée que celle trouvée dans ce travail (6,19 x 10^{-9}), et qui pourrait conduire à un TMRCA moins ancien. La méthode utilisée dans Su et al. (2003) pour calculer le taux de mutation est la même que celle utilisée dans ce travail. Cependant, l'espèce utilisée pour avoir le temps de divergence est *N. caninum* et non pas *H. hammondi*, et l'estimation de ce temps de divergence est de 12,7 Ma donc pas très différent de 11Ma. Or, la divergence entre *N. caninum* et *T. gondii* a déjà été estimée à environ 28 Ma (Reid et al., 2012). Les différences génétiques entre *N. caninum* et *T. gondii* sont sans doute plus grandes qu'entre *H. hammondi* et *T. gondii*. La prise en compte d'un temps divergence similaire dans les 2 cas a conduit à un taux de mutation plus élevé dans l'étude de Su et al. (2003) et donc à un TMRCA plus récent. Les difficultés à estimer le taux de mutations peuvent donc provenir de discordances concernant l'histoire évolutive des

BY-NC-ND

Apicomplexes soulignant l'importance des études phylogénétiques qui permettent d'avoir des informations sur les parasites proches de celui étudié, même si leur importance en santé humaine ou animale est plus limitée.

La date calculée ici correspond à celle du plus récent ancêtre commun des souches de *T. gondii*. Il serait apparu il y a environ 1,5 Ma. Ce résultat est éloigné de la date de divergence entre *H. hammondi* et *T. gondii*, il y a environ 11 Ma. Cet intervalle de temps (1,5–11 Ma) correspondrait à la durée nécessaire pour que la séparation de la lignée se réalise et que les polymorphismes ancestraux soient triés. Dans le cas d'une divergence ancienne, comme c'est le cas ici, une séparation claire existe entre la divergence et le TMRCA. D'après certaines études, la séparation complète de la lignée est attendue à environ 4-7 Ne générations après l'isolation reproductive. Dans le cas d'une divergence récente, il est possible que le tri des lignées ne soit pas complet et qu'il soit difficile de distinguer le TMRCA de la date de divergence. Dans ce type d'études, l'échantillonnage est très important, car l'estimation de la date d'apparition d'un parasite est dépendante de l'espèce sœur la plus proche connue (Pettengill, 2015).

II.6. Discussion générale

Les résultats précédents montrent que le TMRCA de *T. gondii* daterait d'environ 1,5 Ma, et que l'origine des souches actuelles se situerait en Amérique du Sud. Le résultat des datations est cohérent avec celui trouvé dans Su et al., (2003) tandis que l'origine géographique déterminée à l'aide de l'analyse spatiale conforte l'hypothèse émise par Lehmann et al., (2006). Cependant dans notre étude et contrairement à l'étude de Lehmann, l'Amérique du Sud serait l'origine des lignées actuelles de *T. gondii* et non l'origine de l'espèce dont la date d'apparition se situerait aux alentours des 11 Ma. Dans leur étude Lehmann et al. émettent l'hypothèse que seuls les félidés sauvages en faible abondance servaient d'hôtes définitifs au toxoplasme avant l'arrivée du chat domestique au 16ème siècle. Cependant, l'origine de *T. gondii* est plus ancienne que l'arrivée des félidés en Amérique du Sud qui a eu lieu il y a environ 2-3 Ma lors de la reconnexion de l'Isthme de Panama (O'Brien et al., 2008). Ainsi l'hypothèse d'une origine géographique en Amérique du Sud de *T. gondii* impliquerait que ce parasite possédait d'autres hôtes définitifs que les félidés ce qui suppose qu'un changement d'hôte aurait ensuite eu lieu. Lorsque l'Amérique du Sud était isolée de l'Amérique du Nord, aucun carnivore placentaire n'était présent. Il y avait plusieurs espèces de marsupiaux dont les niches écologiques ont été par la suite envahies par les nouveaux félidés (O'Brien et al., 2008). Il serait donc possible qu'un parasite se soit adapté à ce nouvel hôte. Néanmoins, le parasite le plus proche phylogénétiquement de *T. gondii* est *H. hammondi* qui possède également les félidés comme hôtes définitfs. L'hypothèse précédente impliquerait donc que ces deux espèces se soient séparées sur d'autres hôtes que les félidés et qu'elles aient ensuite changé d'hôtes définitifs pour les félidés. Cette hypothèse ne nous a pas semblé être la plus probable et nous avons donc privilégié celle impliquant une origine géographique de l'espèce différente de celle de l'ancêtre commun le plus récent des lignées actuelles. Nous proposons donc que *T. gondii* et *H. hammondi* aient commencé à diverger sur les félidés il y a environ 11 Ma. La séparation en deux espèces différentes est un processus long (Pettengill, 2015). Entre 11 Ma et 1,5 Ma, *T. gondii* était vraisemblablement encore peu différencié d'*H. hammondi* et les parasites dépendaient des migrations des félidés pour coloniser des régions où ils étaient absents. Lors de l'entrée des félidés en Amérique du Sud à partir de l'Amérique du Nord entre 2 et 3 Ma, un *T. gondii* ancestral a pu également y pénétrer. L'échange des faunes entre Amérique du Sud et Amérique du Nord est un processus difficile à comprendre

BY-NC-ND

pour lequel il manque à l'heure actuelle beaucoup d'informations (Carrillo et al., 2015). Des études suggèrent cependant qu'il est probable que les félidés aient trouvé un environnement favorable qui aurait favorisé leur diversification, en particulier celle de la lignée de l'ocelot (ocelot, margay, chat des Andes, chat des pampas, chat de Geoffroy, kodkod et chat tigre). Mais d'autres familles auraient également profité de cet environnement nouveau. Plusieurs genres sont ainsi apparus dans la famille des Muridae, hôtes intermédiaires de *T. gondii* (Webb, 2006). L'hypothèse a été émise que l'ouverture de l'Amérique du Sud et l'échange des faunes entre Amérique du Sud et Amérique du Nord ont pu favoriser une évolution rapide et innovante (Webb, 2006). Dans ces conditions, il est possible que la divergence entre *T. gondii* et *H. hammondi* se soit accélérée. Une différence majeure entre ces deux parasites est l'infectivité des kystes tissulaires de *T. gondii* entre hôtes intermédiaires par carnivorisme alors que les kystes tissulaires des hôtes intermédiaires *d'H. hammondi* ne sont infectants que pour les hôtes définitifs. Cette capacité de transmission par carnivorisme a pu apparaître chez *T. gondii* dans le contexte favorable de la forêt amazonienne avec son grand nombre et sa grande diversité d'hôtes intermédiaires. En Amérique du Sud, les niches écologiques sont variées et l'environnement hétérogène, ce qui peut également expliquer la grande diversité génétique de *T. gondii* dans cette région du monde. La diversification du parasite pourrait en effet être une conséquence de la diversification de ses hôtes intermédiaires. L'acquisition du carnivorisme comme nouveau mode de transmission est dans tous les cas un avantage pour *T. gondii* qui peut se transmettre d'hôte intermédiaire en hôte intermédiaire même dans des milieux où les félidés sont peu présents. Cependant, la dissémination du parasite via les oocystes excrétés par les félidés reste un moyen de transmission très important. L'avantage procuré par cette nouvelle voie de transmission par carnivorisme a pu permettre à *T. gondii* de dominer la forme ancestrale encore proche d'*H. hammondi* et de la remplacer. A travers les déplacements de ses différents hôtes, le toxoplasme « moderne » a pu ensuite remonter en Amérique du Nord. Seuls quelques génotypes de la population de départ d'Amérique du Sud seraient impliqués dans ces migrations. D'après les voies de dissémination représentées dans la figure 26, le parasite aurait ensuite franchi le détroit de Béring puis colonisé l'Asie, l'Afrique et l'Europe. Il se peut donc que la forme ancestrale ait été présente dans ces différentes régions avant l'arrivée de la forme actuelle qui l'aurait supplantée. Les étapes successives de migrations ont créé des effets de fondation qui ont conduit à une perte de diversité progressive pouvant expliquer la structure de population actuelle avec une diversité plus importante en Amérique du Sud, intermédiaire en Amérique du Nord et plus faible en Europe, Asie et Afrique. L'analyse spatiale a montré en effet qu'une partie de la diversité actuelle était expliquée par les distances à l'origine. Des évènements plus récents ont pu cependant modifier la structure originale. Une des caractéristiques de *T. gondii* est la prédominance de types clonaux en Europe, Afrique et Asie. La domestication du chat il y a environ 10 000 ans en Asie du Sud-Ouest puis son expansion d'abord dans l'Ancien Monde a pu agir comme une force de sélection, en amplifiant les génotypes qui étaient déjà les mieux adaptés (Müller and Howard, 2016), permettant leur expansion clonale. Les transports longue-distance des Hommes et des animaux a pu également modifier la structure de départ en permettant le brassage de souches qui étaient géographiquement séparées. Le Gabon présente par exemple des similarités avec certaines souches du Brésil, laissant supposer que les échanges maritimes durant les 18ème et 19ème siècles a pu contribuer à l'apport de souches sud-américaines en Afrique (Mercier et al., 2010). Le scénario proposé ici, en différenciant l'apparition de l'espèce *T. gondii,* ancienne et l'apparition du TMRCA des souches actuelles, beaucoup plus récente, réconcilie les deux études qui semblaient avoir des hypothèses opposées sur l'origine de *T. gondii* (Khan et al., 2007; Lehmann et al., 2006). L'origine du TMRCA serait en Amérique du Sud, en accord avec

BY-NC-ND

Lehmann et al. (2006), mais l'espèce serait sans doute apparue ailleurs, et aurait migré en Amérique du Sud à partir de l'Amérique du Nord lors de la reconnexion de l'Isthme de Panama, scénario favorisé par Khan et al. (2007). Cependant la dissémination du parasite proposé par Lehmann et al. (2006) diffère de celle proposée ici. Lehmann et al. (2006) propose en effet une dissémination assez récente principalement due aux échanges et aux transports humains. Ainsi pendant une grande partie de son histoire, le toxoplasme n'aurait pas été aussi ubiquiste qu'aujourd'hui. Un évènement rare aurait permis la migration du toxoplasme en Eurasie, où il aurait évolué de façon indépendante de la population sud-américaine. Le trafic maritime aurait ensuite permis à cette population de s'introduire en Amérique du Nord et en Asie. Bien qu'en accord sur l'origine géographique du parasite, les deux scénarios sont donc différents pour expliquer les voies de dissémination et la structure actuelle de la population de *T. gondii*. L'origine de l'espèce *T. gondii* est inconnue, et nos données ne nous permettent pas de conclure. L'hypothèse faite ici tendrait cependant à supposer que l'Asie pourrait être une origine probable puisqu'il s'agit de l'origine des félidés modernes qui y seraient apparus il y a environ 10,8 Ma, ce qui correspond approximativement à la date d'apparition de *T. gondii.*

En plus d'être l'origine des souches actuelles de *T. gondii*, l'Amérique du Sud paye le plus lourd tribut à la toxoplasmose. La prévalence, l'incidence et la sévérité des toxoplasmoses oculaires acquises et congénitales dans certaines régions du Brésil, de Colombie et d'Argentine sont considérablement plus élevées que n'importe où ailleurs, ce qui fait de la toxoplasmose oculaire un réel problème de santé publique dans ces régions (Glasner et al., 1992; Gilbert et al., 2008; de-la-Torre et al., 2009; Rudzinski et al., 2016). Parce que l'Amérique du Sud est aussi un point chaud de la diversité génétique de *T. gondii*, il a été émis l'hypothèse que les formes sévères de toxoplasmose pourraient être la conséquence d'une mauvaise adaptation de l'hôte humain à la diversité inhabituelle des souches dans cette partie du monde (Demar et al., 2012; Gilbert et al., 2008; Khan et al., 2006a; Rudzinski et al., 2016; de-la-Torre et al., 2013). En fait, des hôtes d'espèces différentes peuvent avoir une susceptibilité très variable face à une même souche de *T. gondii* et il y a de grandes différences entre les souches dans la modulation des gènes de l'hôte impliqués dans l'activation des réponses inflammatoires (Melo). Ceci suggère que le degré de co-évolution hôte-pathogène joue un rôle majeur dans la physiopathologie de la toxoplasmose. Une réponse immunitaire inappropriée contre certaines souches pourrait ainsi conduire à une forme de la maladie plus agressive (de la Torre et al., 2013). Notre analyse phylogéographique indique que les souches sauvages de la forêt amazonienne sont les plus ancestrales et sont les vestiges de la diversité originelle qui est apparue à cet endroit il y a 1,5 Ma. Elles entrainent des formes sévères même chez l'immunocompétent car la majeure partie de l'humanité a vécu loin de ces souches confinées dans cet environnement dépeuplé. Inversement, du fait d'une constante co-évolution hôte-pathogène à travers le cycle domestique de *T. gondii*, les humains seraient mieux adaptés aux génotypes les plus récents qui ont émergé il y a environ 10 000 ans dans le contexte de la domestication animale et de l'anthropisation de l'environnement (Mercier et al., 2011; Su et al., 2003). L'exemple le plus parlant d'adaptation hôte-pathogène réussie est donné par les souches clonales de type II qui représentent plus de 90% des souches animales et humaines en Europe et qui est considérablement moins pathogène pour l'œil que les souches sud-américaines (Ajzenberg, 2015; de-la-Torre et al., 2013). La coévolution ou son absence avec certains génotypes de *T. gondii* serait donc un déterminant majeur de la sévérité de la toxoplasmose humaine, en particulier en Amérique du Sud le lieu de naissance da la forme moderne de ce parasite cosmopolite.

BY-NC-ND

Perspectives et conclusion

Les analyses pourraient bénéficier d'un échantillonnage plus diversifié, en particulier en Afrique, en Asie et en Océanie. Des isolats provenant d'hôtes sauvages devraient également être pris en compte dans les populations non-Américaines. Les biais d'échantillonnage semblent inévitables et des analyses plus approfondies prenant en compte un éventuel lien entre les diversités génétiques des populations et la diversité des hôtes considérés ou la nature des hôtes (intermédiaires ou définitifs) pourraient permettre d'avoir des résultats plus robustes. Les marqueurs pourraient aussi être changés, en prenant de préférence des marqueurs plus neutres afin d'éviter les effets de la sélection ou provenant des gènes de génome de l'apicoplaste ou de mitochondrie pour limiter les recombinaisons. Cela permettrait de conforter les résultats décrits, comme cela a été fait pour *Plasmodium falciparum* (Tanabe et al., 2010, 2013). Toujours en ce qui concerne les marqueurs, l'analyse préliminaire effectuée à partir des séquences présentes sur GenBank pourrait être appronfondie. Il s'agit en effet d'une source importante d'information et certains loci qui ont été séquencés pour un certain nombre de taxons identiques pourraient être utilisés pour identifier des loci et/ou des taxons présentant une histoire évolutive particulière. Ces alignements pourraient aussi servir pour conforter les résultats trouvés sur les taux de mutations et le TMRCA si les séquences correspondantes étaient retrouvées chez *H. hammondi*. Il serait également intéressant de pouvoir dater les évènements de migration d'une région à l'autre afin de pouvoir différencier les évènements appartenant à l'histoire ancienne de ceux plus récents, et ainsi départager les différents scenarii proposés sur les voies de dissémination du parasite. Le manque de connaissances sur la phylogénie et la biologie des Apicomplexes, essentiellement dû à un échantillonnage opportuniste a déjà été souligné (Morrison, 2009). *H. hammondi* est le parasite le plus proche connu de *T. gondii*. Il est donc celui qui est susceptible d'apporter le plus d'informations sur son évolution. Malheureusement il est très peu étudié. Bien que son cycle soit connu, tous ses hôtes naturels n'ont peut-être pas été déterminés et très peu d'informations sont disponibles sur sa distribution. Il serait donc intéressant, d'étudier en parallèle les deux parasites. Avoir plus d'informations sur la distribution géographique et la diversité génétique d'*H. hammondi* pourrait permettre de voir s'il possède les mêmes caractéristiques que *T. gondii*. Cela serait utile pour formuler des hypothèses, notamment sur les conséquences des différences du cycle de vie. Des analyses comparatives de génomes pourraient aussi permettre d'élucider les bases génétiques de la différence de capacité de transmission via les kystes entre les hôtes intermédiaires.

Les coûts sociaux et économiques de la prise en charge des cas symptomatiques de toxoplasmose congénitale peuvent être considérables mais le rapport coût-efficacité du dépistage prénatal ou néonatal de la maladie et du traitement des enfants atteints est encore débattu (Jones et al., 2014; Wallon et al., 1999). Il y a un réel besoin d'essais cliniques randomisés et contrôlés par placebo pour aider à déterminer l'efficacité de ces interventions. Notre reconstruction de la phylogéographie de *T. gondii* fournit un modèle pour la compréhension des profils de diversité génétique dans les endroits où le parasite a été échantillonné et pour la prédiction de la diversité dans les localités non échantillonnées. Parce que la diversité génétique du parasite semble être en lien la sévérité de la toxoplasmose, nos résultats peuvent être utilisés comme une base pour expliquer les différences géographiques dans le fardeau de la maladie et pour identifier les endroits prioritaires pour y réaliser des essais cliniques dans le futur.

BY-NC-ND

Références bibliographiques

Adl, S.M., Leander, B.S., Simpson, A.G.B., Archibald, J.M., Anderson, O.R., Bass, D., Bowser, S.S., Brugerolle, G., Farmer, M.A., Karpov, S., et al. (2007). Diversity, Nomenclature, and Taxonomy of Protists. Syst. Biol. *56*, 684–689.

Ajzenberg, D. (2012). High burden of congenital toxoplasmosis in the United States: the strain hypothesis? Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. *54*, 1606–1607.

Ajzenberg, D. (2015). 1995-2015: it is time to celebrate 20 years of (intensive) genotyping of *Toxoplasma gondii* strains. Future Microbiol. *10*, 689–691.

Ajzenberg, D., Cogné, N., Paris, L., Bessières, M.-H., Thulliez, P., Filisetti, D., Pelloux, H., Marty, P., and Dardé, M.-L. (2002). Genotype of 86 *Toxoplasma gondii* Isolates Associated with Human Congenital Toxoplasmosis, and Correlation with Clinical Findings. J. Infect. Dis. *186*, 684–689.

Ajzenberg, D., Banuls, A.L., Su, C., Dumetre, A., Demar, M., Carme, B., and Darde, M.L. (2004). Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. Int. J. Parasitol. *34*, 1185–1196.

Ajzenberg, D., Carme, B., Demar, M., Boukhari, R., and Dardé, M.-L. (2007). La toxoplasmose ≪ guyanaise ≫. Rev. Francoph. Lab. *2007*, 51–60.

Ajzenberg, D., Collinet, F., Mercier, A., Vignoles, P., and Dardé, M.-L. (2010). Genotyping of *Toxoplasma gondii* Isolates with 15 Microsatellite Markers in a Single Multiplex PCR Assay. J. Clin. Microbiol. *48*, 4641–4645.

Alaganan, A., Fentress, S.J., Tang, K., Wang, Q., and Sibley, L.D. (2014). Toxoplasma GRA7 effector increases turnover of immunity-related GTPases and contributes to acute virulence in the mouse. Proc. Natl. Acad. Sci. U. S. A. *111*, 1126–1131.

Andenmatten, N., Egarter, S., Jackson, A.J., Jullien, N., Herman, J.-P., and Meissner, M. (2013). Conditional genome engineering in *Toxoplasma gondii* uncovers alternative invasion mechanisms. Nat. Methods *10*, 125–127.

Arisue, N., and Hashimoto, T. (2015). Phylogeny and evolution of apicoplasts and apicomplexan parasites. Parasitol. Int. *64*, 254–259.

Bischoff, A.R., Friedrich, L., Cattan, J.M., and de Freitas Uberti, F.A. (2016). Incidence of Symptomatic Congenital Toxoplasmosis During 10 Years in a Brazilian Hospital. Pediatr. Infect. Dis. J.

Blader, I.J., Coleman, B.I., Chen, C.-T., and Gubbels, M.-J. (2015). Lytic Cycle of *Toxoplasma gondii*: 15 Years Later. Annu. Rev. Microbiol. *69*, 463–485.

Bontell, I.L., Hall, N., Ashelford, K.E., Dubey, J., Boyle, J.P., Lindh, J., and Smith, J.E. (2009). Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants. Genome Biol. *10*, R53.

Boyle, J.P., Rajasekar, B., Saeij, J.P.J., Ajioka, J.W., Berriman, M., Paulsen, I., Roos, D.S., Sibley, L.D., White, M.W., and Boothroyd, J.C. (2006). Just one cross appears capable of

BY-NC-ND

dramatically altering the population biology of a eukaryotic pathogen like *Toxoplasma gondii*. Proc. Natl. Acad. Sci. U. S. A. *103*, 10514–10519.

Carme, B., Bissuel, F., Ajzenberg, D., Bouyne, R., Aznar, C., Demar, M., Bichat, S., Louvel, D., Bourbigot, A.M., Peneau, C., et al. (2002). Severe Acquired Toxoplasmosis in Immunocompetent Adult Patients in French Guiana. J. Clin. Microbiol. *40*, 4037–4044.

Carme, B., Demar, M., Ajzenberg, D., and Dardé, M.L. (2009). Severe acquired toxoplasmosis caused by wild cycle of *Toxoplasma gondii*, French Guiana. Emerg. Infect. Dis. *15*, 656–658.

Carrillo, J.D., Forasiepi, A., Jaramillo, C., and Sánchez-Villagra, M.R. (2015). Neotropical mammal diversity and the Great American Biotic Interchange: spatial and temporal variation in South America's fossil record. Evol. Popul. Genet. *5*, 451.

Clark, E.L., Macdonald, S.E., Thenmozhi, V., Kundu, K., Garg, R., Kumar, S., Ayoade, S., Fornace, K.M., Jatau, I.D., Moftah, A., et al. (2016). Cryptic *Eimeria* genotypes are common across the southern but not northern hemisphere. Int. J. Parasitol.

CNR Toxoplasmose (2015). Rapport annuel d'activités (Centre national de référence de la Toxoplasmose).

Darde, M.-L. (1997). Apport des typages de souches de toxoplasmes dans l'épidémiologie de la toxoplasmose. Rev. Fr. Lab. *1997*, 144–148.

Dardé, M.L., Bouteille, B., and Pestre-Alexandre, M. (1992). Isoenzyme analysis of 35 *Toxoplasma gondii* isolates and the biological and epidemiological implications. J. Parasitol. *78*, 786–794.

Delhaes, L., Ajzenberg, D., Sicot, B., Bourgeot, P., Dardé, M.-L., Dei-Cas, E., and Houfflin-Debarge, V. (2010). Severe congenital toxoplasmosis due to a *Toxoplasma gondii* strain with an atypical genotype: case report and review. Prenat. Diagn. *30*, 902–905.

Delsuc, F., and Douzery, E.J. (2004). Les méthodes probabilistes en phylogénie moléculaire:(1) Les modèles d'évolution des séquences et le maximum de vraisemblance. Biosystema *22*, 59–74.

Demar, M., Ajzenberg, D., Maubon, D., Djossou, F., Panchoe, D., Punwasi, W., Valery, N., Peneau, C., Daigre, J.-L., Aznar, C., et al. (2007). Fatal Outbreak of Human Toxoplasmosis along the Maroni River: Epidemiological, Clinical, and Parasitological Aspects. Clin. Infect. Dis. *45*, e88–e95.

Demar, M., Hommel, D., Djossou, F., Peneau, C., Boukhari, R., Louvel, D., Bourbigot, A.-M., Nasser, V., Ajzenberg, D., Darde, M.-L., et al. (2012). Acute toxoplasmoses in immunocompetent patients hospitalized in an intensive care unit in French Guiana. Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis. *18*, E221-231.

Desper, R., and Gascuel, O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J. Comput. Biol. J. Comput. Mol. Cell Biol. *9*, 687–705.

Desper, R., and Gascuel, O. (2004). Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. Mol. Biol. Evol. *21*, 587–598.

Douzery, E.J.P., Delsuc, F., and Philippe, H. (2006). Les datations moléculaires à l'heure de la génomique. Médecine/Sciences *22*, 374–380.

BY-NC-ND

Drewry, L.L., and Sibley, L.D. (2015). *Toxoplasma* Actin Is Required for Efficient Host Cell Invasion. mBio *6*.

Dubey, J.P. (2001). Oocyst shedding by cats fed isolated bradyzoites and comparison of infectivity of bradyzoites of the VEG strain *Toxoplasma gondii* to cats and mice. J. Parasitol. *87*, 215–219.

Dubey, J.., and Sreekumar, C. (2003). Redescription of *Hammondia hammondi* and its differentiation from *Toxoplasma gondii*. Int. J. Parasitol. *33*, 1437–1453.

Dubey, J.P., Huong, L.T.T., Sundar, N., and Su, C. (2007b). Genetic characterization of *Toxoplasma gondii* isolates in dogs from Vietnam suggests their South American origin. Vet. Parasitol. *146*, 347–351.

Dubey, J.P., Rajapakse, R.P.V.J., Wijesundera, R.R.M.K.K., Sundar, N., Velmurugan, G.V., Kwok, O.C.H., and Su, C. (2007a). Prevalence of *Toxoplasma gondii* in dogs from Sri Lanka and genetic characterization of the parasite isolates. Vet. Parasitol. *146*, 341–346.

Dubey, J.P., Lago, E.G., Gennari, S.M., Su, C., and Jones, J.L. (2012). Toxoplasmosis in humans and animals in Brazil: high prevalence, high burden of disease, and epidemiology. Parasitology *139*, 1375–1424.

Dubey, J.P., Choudhary, S., Tilahun, G., Tiao, N., Gebreyes, W.A., Zou, X., and Su, C. (2013). Molecular and Biological Characterization of First Isolates of *Hammondia hammondi* from Cats from Ethiopia. J. Parasitol. *99*.

Dubremetz, J.F., and Lebrun, M. (2012). Virulence factors of *Toxoplasma gondii*. Microbes Infect. Inst. Pasteur *14*, 1403–1410.

E Keats Shwab, X.-Q.Z. (2013). Geographical patterns of *Toxoplasma gondii* genetic diversity revealed by multilocus PCR-RFLP genotyping. Parasitology *141*, 1–9.

Faria, N.R., Rambaut, A., Suchard, M.A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pépin, J., et al. (2014). The early spread and epidemic ignition of HIV-1 in human populations. Science *346*, 56–61.

Fazaeli, A., and Ebrahimzadeh, A. (2007). A new perspective on and re-assessment of SAG2 locus as the tool for genetic analysis of *Toxoplasma gondii* isolates. Parasitol. Res. *101*, 99–104.

Fazaeli, A., Carter, P.E., Darde, M.L., and Pennington, T.H. (2000). Molecular typing of *Toxoplasma gondii* strains by GRA6 gene sequence analysis. Int. J. Parasitol. *30*, 637–642.

Felsenstein, J. (2004). Inferring Phylogenies (Macmillan Education).

Ferguson, D.J. (2009). *Toxoplasma gondii*: 1908-2008, homage to Nicolle, Manceaux and Splendore. Mem. Inst. Oswaldo Cruz *104*, 133–148.

Ferguson, D.J.P., Henriquez, F.L., Kirisits, M.J., Muench, S.P., Prigge, S.T., Rice, D.W., Roberts, C.W., and McLeod, R.L. (2005). Maternal Inheritance and Stage-Specific Variation of the Apicoplast in *Toxoplasma gondii* during Development in the Intermediate and Definitive Host. Eukaryot. Cell *4*, 814–826.

Fox, B.A., and Bzik, D.J. (2002). De novo pyrimidine biosynthesis is required for virulence of *Toxoplasma gondii*. Nature *415*, 926–929.

BY-NC-ND

Frénal, K., and Soldati-Favre, D. (2013). Un complexe moléculaire unique à l'origine de la motilité et de l'invasion des Apicomplexes. Médecine/Sciences *29*, 515–522.

Frerichs, R.R., Keim, P. s., Barrais, R., and Piarroux, R. (2012). Nepalese origin of cholera epidemic in Haiti. Clin. Microbiol. Infect. *18*, E158–E163.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. *14*, 685–695.

Gilbert, R.E., Freeman, K., Lago, E.G., Bahia-Oliveira, L.M.G., Tan, H.K., Wallon, M., Buffolano, W., Stanford, M.R., and Petersen, E. (2008). Ocular Sequelae of Congenital Toxoplasmosis in Brazil Compared with Europe. PLoS Negl. Trop. Dis. *2*.

Glasner, P.D., Silveira, C., Kruszon-Moran, D., Martins, M.C., Burnier Júnior, M., Silveira, S., Camargo, M.E., Nussenblatt, R.B., Kaslow, R.A., and Belfort Júnior, R. (1992). An unusually high prevalence of ocular toxoplasmosis in southern Brazil. Am. J. Ophthalmol. *114*, 136–144.

Gómez-Marin, J.E., de-la-Torre, A., Angel-Muller, E., Rubio, J., Arenas, J., Osorio, E., Nuñez, L., Pinzon, L., Mendez-Cordoba, L.C., Bustos, A., et al. (2011). First Colombian Multicentric Newborn Screening for Congenital Toxoplasmosis. PLoS Negl. Trop. Dis. *5*.

Gómez-Marín, J.E., de-la-Torre, A., Barrios, P., Cardona, N., Álvarez, C., and Herrera, C. (2012). Toxoplasmosis in military personnel involved in jungle operations. Acta Trop. *122*, 46–51.

Grigg, M.E., and Sundar, N. (2009). Sexual recombination punctuated by outbreaks and clonal expansions predicts *Toxoplasma gondii* population genetics. Int. J. Parasitol. *39*, 925–933.

Grigg, M.E., and Suzuki, Y. (2003). Sexual recombination and clonal evolution of virulence in *Toxoplasma*. Microbes Infect. *5*, 685–690.

Grigg, M.E., Bonnefoy, S., Hehl, A.B., Suzuki, Y., and Boothroyd, J.C. (2001). Success and virulence in *toxoplasma* as the result of sexual recombination between two distinct ancestries. Science *294*, 161–165.

Groh, M., Faussart, A., Villena, I., Ajzenberg, D., Carme, B., Demar, M., Joly, V., Houze, S., Simon, S., Aubert, D., et al. (2012). Acute lung, heart, liver, and pancreatic involvements with hyponatremia and retinochoroiditis in a 33-year-old French Guianan patient. PLoS Negl. Trop. Dis. *6*, e1802.

Guégan, J.-F., and Choisy, M. (2009). Introduction à l'épidémiologie intégrative des maladies infectieuses et parasitaires.

Hahn, B.H., Shaw, G.M., De Cock, K.M., and Sharp, P.M. (2000). AIDS as a zoonosis: scientific and public health implications. Science *287*, 607–614.

Han, B.A., Kramer, A.M., and Drake, J.M. (2016). Global Patterns of Zoonotic Disease in Mammals. Trends Parasitol. *32*, 565–577.

Holland, G.N. (2003). Ocular toxoplasmosis: a global reassessment. Part I: epidemiology and course of disease. Am. J. Ophthalmol. *136*, 973–988.

Howe, D.K., and Sibley, L.D. (1994). *Toxoplasma gondii*: analysis of different laboratory stocks of the RH strain reveals genetic heterogeneity. Exp. Parasitol. *78*, 242–245.

BY-NC-ND

Howe, D.K., Honoré, S., Derouin, F., and Sibley, L.D. (1997). Determination of genotypes of *Toxoplasma gondii* strains isolated from patients with toxoplasmosis. J. Clin. Microbiol. *35*, 1411–1414.

Hunter, C.A., and Sibley, L.D. (2012). Modulation of innate immunity by *Toxoplasma gondii* virulence effectors. Nat. Rev. Microbiol. *10*, 766–778.

Iltzsch, M.H. (1993). Pyrimidine Salvage Pathways In *Toxoplasma gondii*. J. Eukaryot. Microbiol. *40*, 24–28.

Jamieson, S.E., de Roubaix, L.-A., Cortina-Borja, M., Tan, H.K., Mui, E.J., Cordell, H.J., Kirisits, M.J., Miller, E.N., Peacock, C.S., Hargrave, A.C., et al. (2008). Genetic and epigenetic factors at COL2A1 and ABCA4 influence clinical outcome in congenital toxoplasmosis. PloS One *3*, e2285.

Johnson, W.E. (2006). The Late Miocene Radiation of Modern Felidae: A Genetic Assessment. Science *311*, 73–77.

Johnson, W.E., and O'Brien, S.J. (1997). Phylogenetic reconstruction of the Felidae using 16S rRNA and NADH-5 mitochondrial genes. J. Mol. Evol. *44 Suppl 1*, S98-116.

Jones, J.L., Parise, M.E., and Fiore, A.E. (2014). Neglected parasitic infections in the United States: toxoplasmosis. Am. J. Trop. Med. Hyg. *90*, 794–799.

Khan, A., Jordan, C., Muccioli, C., Vallochi, A.L., Rizzo, L.V., Belfort, R., Vitor, R.W.A., Silveira, C., and Sibley, L.D. (2006a). Genetic divergence of *Toxoplasma gondii* strains associated with ocular toxoplasmosis, Brazil. Emerg. Infect. Dis. *12*, 942–949.

Khan, A., Böhme, U., Kelly, K.A., Adlem, E., Brooks, K., Simmonds, M., Mungall, K., Quail, M.A., Arrowsmith, C., Chillingworth, T., et al. (2006b). Common inheritance of chromosome Ia associated with clonal expansion of *Toxoplasma gondii*. Genome Res. *16*, 1119–1125.

Khan, A., Fux, B., Su, C., Dubey, J.P., Darde, M.L., Ajioka, J.W., Rosenthal, B.M., and Sibley, L.D. (2007). Recent transcontinental sweep of *Toxoplasma gondii* driven by a single monomorphic chromosome. Proc. Natl. Acad. Sci. *104*, 14872–14877.

Khan, A., Taylor, S., Ajioka, J.W., Rosenthal, B.M., and Sibley, L.D. (2009). Selection at a single locus leads to widespread expansion of *Toxoplasma gondii* lineages that are virulent in mice. PLoS Genet. *5*, e1000404.

Khan, A., Dubey, J.P., Su, C., Ajioka, J.W., Rosenthal, B.M., and Sibley, L.D. (2011a). Genetic analyses of atypical *Toxoplasma gondii* strains reveal a fourth clonal lineage in North America. Int. J. Parasitol. *41*, 645–655.

Khan, A., Miller, N., Roos, D.S., Dubey, J.P., Ajzenberg, D., Dardé, M.L., Ajioka, J.W., Rosenthal, B., and Sibley, L.D. (2011b). A monomorphic haplotype of chromosome Ia is associated with widespread success in clonal and nonclonal populations of *Toxoplasma gondii*. mBio *2*, e00228-00211.

Lehmann, T., Blackston, C.R., Parmley, S.F., and Dubey, J.P. (2000). Strain Typing of *Toxoplasma gondii*: Comparison of Anigen-Coding and Housekeeping Genes. J. Parasitol. *86*, 960–971.

Lehmann, T., Marcet, P.L., Graham, D.H., Dahl, E.R., and Dubey, J.P. (2006). Globalization and the population structure of *Toxoplasma gondii*. Proc. Natl. Acad. Sci. *103*, 11423–11428.

BY-NC-ND

Lekutis, C., Ferguson, D.J.P., Grigg, M.E., Camps, M., and Boothroyd, J.C. (2001). Surface antigens of *Toxoplasma gondii*: variations on a theme. Int. J. Parasitol. *31*, 1285–1292.

Lemey, P., Pybus, O.G., Wang, B., Saksena, N.K., Salemi, M., and Vandamme, A.-M. (2003). Tracing the origin and history of the HIV-2 epidemic. Proc. Natl. Acad. Sci. U. S. A. *100*, 6588–6592.

Lopes-Mori, F.M.R., Mitsuka-Breganó, R., Bittencourt, L.H.F. de B., Dias, R.C.F., Gonçalves, D.D., Capobiango, J.D., Reiche, E.M.V., Morimoto, H.K., Freire, R.L., and Navarro, I.T. (2013). Gestational toxoplasmosis in Paraná State, Brazil: prevalence of IgG antibodies and associated risk factors. Braz. J. Infect. Dis. Off. Publ. Braz. Soc. Infect. Dis. *17*, 405–409.

Lorenzi, H., Khan, A., Behnke, M.S., Namasivayam, S., Swapna, L.S., Hadjithomas, M., Karamycheva, S., Pinney, D., Brunk, B.P., Ajioka, J.W., et al. (2016). Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. Nat. Commun. *7*, 10147.

Ma, J.S., Sasai, M., Ohshima, J., Lee, Y., Bando, H., Takeda, K., and Yamamoto, M. (2014). Selective and strain-specific NFAT4 activation by the *Toxoplasma gondii* polymorphic dense granule protein GRA6. J. Exp. Med. *211*, 2013–2032.

McLeod, R., Boyer, K.M., Lee, D., Mui, E., Wroblewski, K., Karrison, T., Noble, A.G., Withers, S., Swisher, C.N., Heydemann, P.T., et al. (2012). Prematurity and Severity Are Associated With *Toxoplasma gondii* Alleles (NCCCTS, 1981–2009). Clin. Infect. Dis. *54*, 1595–1605.

Mercier, A. (2010). Approche écologique, épidémiologique et génétique de la biodiversité de *Toxoplasma gondii* en zone tropicale humide: exemples du Gabon et de la Guyane Française. Université de Limoges.

Mercier, C., and Cesbron-Delauw, M.-F. (2015). *Toxoplasma* secretory granules: one population or more? Trends Parasitol. *31*, 60–71.

Mercier, A., Devillard, S., Ngoubangoye, B., Bonnabau, H., Bañuls, A.-L., Durand, P., Salle, B., Ajzenberg, D., and Dardé, M.-L. (2010). Additional Haplogroups of *Toxoplasma gondii* out of Africa: Population Structure and Mouse-Virulence of Strains from Gabon. PLoS Negl. Trop. Dis. *4*, e876.

Mercier, A., Ajzenberg, D., Devillard, S., Demar, M.P., de Thoisy, B., Bonnabau, H., Collinet, F., Boukhari, R., Blanchet, D., Simon, S., et al. (2011). Human impact on genetic diversity of *Toxoplasma gondii*: Example of the anthropized environment from French Guiana. Infect. Genet. Evol. *11*, 1378–1387.

Mercier, C., Dubremetz, J.-F., Rauscher, B., Lecordier, L., Sibley, L.D., and Cesbron-Delauw, M.-F. (2002). Biogenesis of Nanotubular Network in *Toxoplasma* Parasitophorous Vacuole Induced by Parasite Proteins. Mol. Biol. Cell *13*, 2397–2409.

Millar, P.R., Moura, F.L. de, Bastos, O.M.P., Mattos, D.P.B.G. de, Fonseca, A.B.M., Sudré, A.P., Leles, D., and Amendoeira, M.R.R. (2014). Toxoplasmosis-related knowledge among pregnant and postpartum women attended in public health units in Niterói, Rio De Janeiro, Brazil. Rev. Inst. Med. Trop. São Paulo *56*, 433–438.

Miller, M.., Grigg, M.., Kreuder, C., James, E.., Melli, A.., Crosbie, P.., Jessup, D.., Boothroyd, J.., Brownstein, D., and Conrad, P.. (2004). An unusual genotype of *Toxoplasma gondii* is common in California sea otters (Enhydra lutris nereis) and is a cause of mortality. Int. J. Parasitol. *34*, 275–284.

BY-NC-ND

Minot, S., Melo, M.B., Li, F., Lu, D., Niedelman, W., Levine, S.S., and Saeij, J.P.J. (2012). Admixture and recombination among *Toxoplasma gondii* lineages explain global genome diversity. Proc. Natl. Acad. Sci. U. S. A. *109*, 13458–13463.

Mita, T., and Jombart, T. (2015). Patterns and dynamics of genetic diversity in *Plasmodium falciparum*: What past human migrations tell us about malaria. Parasitol. Int. *64*, 238–243.

Monot, M., Honoré, N., Garnier, T., Araoz, R., Coppée, J.-Y., Lacroix, C., Sow, S., Spencer, J.S., Truman, R.W., Williams, D.L., et al. (2005). On the Origin of Leprosy. Science *308*, 1040–1042.

Monot, M., Honoré, N., Garnier, T., Zidane, N., Sherafi, D., Paniz-Mondolfi, A., Matsuoka, M., Taylor, G.M., Donoghue, H.D., Bouwman, A., et al. (2009). Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat. Genet. *41*, 1282–1289.

Morrison, D.A. (2005). How old are the extant lineages of *Toxoplasma gondii*? Parassitologia *47*, 205–214.

Morrison, D.A. (2009). Evolution of the Apicomplexa: where are we now? Trends Parasitol. *25*, 375–382.

Morrison, D.A., Bornstein, S., Thebo, P., Wernery, U., Kinne, J., and Mattsson, J.G. (2004). The current status of the small subunit rRNA phylogeny of the coccidia (Sporozoa). Int. J. Parasitol. *34*, 501–514.

Morrissette, N. (2015). Targeting Toxoplasma Tubules: Tubulin, Microtubules, and Associated Proteins in a Human Pathogen. Eukaryot. Cell *14*, 2–12.

Müller, U.B., and Howard, J.C. (2016). The impact of *Toxoplasma gondii* on the mammalian genome. Curr. Opin. Microbiol. *32*, 19–25.

Nene, V., Kiara, H., Lacasta, A., Pelle, R., Svitek, N., and Steinaa, L. (2016). The biology of *Theileria parva* and control of East Coast fever – Current status and future trends. Ticks Tick-Borne Dis. *7*, 549–564.

Ng, H.C., Singh, M., and Jeyaseelan, K. (1997). Nucleotide sequence of ToxPK1 gene from *Toxoplasma gondii*. DNA Seq. J. DNA Seq. Mapp. *7*, 179–191.

O'Brien, S.J., and Johnson, W.E. (2007). The evolution of cats. Genomic paw prints in the DNA of the world's wild cats have clarified the cat family tree and uncovered several remarkable migrations in their past. Sci. Am. *297*, 68–75.

O'Brien, S.J., Johnson, W., Driscoll, C., Pontius, J., Pecon-Slattery, J., and Menotti-Raymond, M. (2008). State of cat genomics. Trends Genet. *24*, 268–279.

Pappas, G., Roussos, N., and Falagas, M.E. (2009). Toxoplasmosis snapshots: global status of *Toxoplasma gondii* seroprevalence and implications for pregnancy and congenital toxoplasmosis. Int. J. Parasitol. *39*, 1385–1394.

Patton, S., Rabinowitz, A., Randolph, S., and Johnson, S.S. (1986). A coprological survey of parasites of wild neotropical felidae. J. Parasitol. *72*, 517–520.

Pecon Slattery, J., and O'Brien, S.J. (1998). Patterns of Y and X chromosome DNA sequence divergence during the Felidae radiation. Genetics *148*, 1245–1255.

BY-NC-ND

Pecon-Slattery, J., Pearks Wilkerson, A.J., Murphy, W.J., and O'Brien, S.J. (2004). Phylogenetic assessment of introns and SINEs within the Y chromosome using the cat family felidae as a species tree. Mol. Biol. Evol. *21*, 2299–2309.

Pena, H.F.J., Gennari, S.M., Dubey, J.P., and Su, C. (2008). Population structure and mouse-virulence of *Toxoplasma gondii* in Brazil. Int. J. Parasitol. *38*, 561–569.

Pettengill, J.B. (2015). The Time to Most Recent Common Ancestor Does Not (Usually) Approximate the Date of Divergence. PLOS ONE *10*, e0128407.

Poulin, R., and Morand, S. (2000). The Diversity of Parasites. Q. Rev. Biol. *75*, 277–293.

Ramírez, J.D., and Llewellyn, M.S. (2014). Reproductive clonality in protozoan pathogens--truth or artefact? Mol. Ecol. *23*, 4195–4202.

Reid, A.J., Vermont, S.J., Cotton, J.A., Harris, D., Hill-Cawthorne, G.A., Könen-Waisman, S., Latham, S.M., Mourier, T., Norton, R., Quail, M.A., et al. (2012). Comparative Genomics of the Apicomplexan Parasites *Toxoplasma gondii* and *Neospora caninum*: Coccidia Differing in Host Range and Transmission Strategy. PLoS Pathog. *8*.

Rico-Torres, C.P., Vargas-Villavicencio, J.A., and Correa, D. (2016). Is *Toxoplasma gondii* type related to clinical outcome in human congenital infection? Systematic and critical review. Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol. *35*, 1079–1088.

Rudzinski, M., Khoury, M., Couto, C., and Ajzenberg, D. (2016). Reactivation of Ocular Toxoplasmosis in Non-Hispanic Persons, Misiones Province, Argentina. Emerg. Infect. Dis. *22*, 912–913.

Saeij, J.P.J., Boyle, J.P., and Boothroyd, J.C. (2005). Differences among the three major strains of *Toxoplasma gondii* and their specific interactions with the infected host. Trends Parasitol. *21*, 476–481.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Schemarova, I.V. (2006). The role of tyrosine phosphorylation in regulation of signal transduction pathways in unicellular eukaryotes. Curr. Issues Mol. Biol. *8*, 27–49.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS Pandemic. Cold Spring Harb. Perspect. Med. *1*.

Sibley, L.D., and Ajioka, J.W. (2008). Population Structure of *Toxoplasma gondii*: Clonal Expansion Driven by Infrequent Recombination and Selective Sweeps. Annu. Rev. Microbiol. *62*, 329–351.

Smith, K.F., and Guégan, J.-F. (2010). Changing Geographic Distributions of Human Pathogens. Annu. Rev. Ecol. Evol. Syst. *41*, 231–250.

Stevens, J.R., Noyes, H.A., Dover, G.A., and Gibson, W.C. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi.* Parasitology *118 (Pt 1)*, 107–116.

Su, C., Evans, D., Cole, R.H., Kissinger, J.C., Ajioka, J.W., and Sibley, L.D. (2003). Recent expansion of *Toxoplasma* through enhanced oral transmission. Science *299*, 414–416.

BY-NC-ND

Su, C., Shwab, E.K., Zhou, P., Zhu, X.Q., and Dubey, J.P. (2009). Moving towards an integrated approach to molecular detection and identification of *Toxoplasma gondii*. Parasitology *137*, 1.

Su, C., Khan, A., Zhou, P., Majumdar, D., Ajzenberg, D., Dardé, M.-L., Zhu, X.-Q., Ajioka, J.W., Rosenthal, B.M., Dubey, J.P., et al. (2012). Globally diverse *Toxoplasma gondii* isolates comprise six major clades originating from a small number of distinct ancestral lineages. Proc. Natl. Acad. Sci. U. S. A. *109*, 5844–5849.

Sundar, N., Cole, R.A., Thomas, N.J., Majumdar, D., Dubey, J.P., and Su, C. (2008). Genetic diversity among sea otter isolates of *Toxoplasma gondii*. Vet. Parasitol. *151*, 125–132.

Tanabe, K., Mita, T., Jombart, T., Eriksson, A., Horibe, S., Palacpac, N., Ranford-Cartwright, L., Sawai, H., Sakihama, N., Ohmae, H., et al. (2010). *Plasmodium falciparum* Accompanied the Human Expansion out of Africa. Curr. Biol. *20*, 1283–1289.

Tanabe, K., Mita, T., Palacpac, N.M.Q., Arisue, N., Tougan, T., Kawai, S., Jombart, T., Kobayashi, F., and Horii, T. (2013). Within-population genetic diversity of *Plasmodium falciparum* vaccine candidate antigens reveals geographic distance from a Central sub-Saharan African origin. Vaccine *31*, 1334–1339.

Tenter, A.M., Barta, J.R., Beveridge, I., Duszynski, D.W., Mehlhorn, H., Morrison, D.A., Thompson, R.C.A., and Conrad, P.A. (2002). The conceptual basis for a new classification of the coccidia. Int. J. Parasitol. *32*, 595–616.

Thomas, F., Lefevre, T., and Raymond, M. (2010). Biologie évolutive (De Boeck Université).

Tibayrenc, M., and Ayala, F.J. (2014). New insights into clonality and panmixia in *Plasmodium* and *toxoplasma*. Adv. Parasitol. *84*, 253–268.

Torgerson, P.R., and Mastroiacovo, P. (2013). The global burden of congenital toxoplasmosis: a systematic review. Bull. World Health Organ. *91*, 501–508.

de-la-Torre, A., López-Castillo, C.A., and Gómez-Marín, J.E. (2009). Incidence and clinical characteristics in a Colombian cohort of ocular toxoplasmosis. Eye Lond. Engl. *23*, 1090–1093.

de-la-Torre, A., Sauer, A., Pfaff, A.W., Bourcier, T., Brunet, J., Speeg-Schatz, C., Ballonzoli, L., Villard, O., Ajzenberg, D., Sundar, N., et al. (2013). Severe South American ocular toxoplasmosis is associated with decreased Ifn-γ/Il-17a and increased Il-6/Il-13 intraocular levels. PLoS Negl. Trop. Dis. *7*, e2541.

Vasconcelos-Santos, D.V., Machado Azevedo, D.O., Campos, W.R., Oréfice, F., Queiroz-Andrade, G.M., Carellos, E.V.M., Castro Romanelli, R.M., Januário, J.N., Resende, L.M., Martins-Filho, O.A., et al. (2009). Congenital toxoplasmosis in southeastern Brazil: results of early ophthalmologic examination of a large cohort of neonates. Ophthalmology *116*, 2199–2205.e1.

Vitaliano, S.N., Soares, H.S., Minervino, A.H.H., Santos, A.L.Q., Werther, K., Marvulo, M.F.V., Siqueira, D.B., Pena, H.F.J., Soares, R.M., Su, C., et al. (2014). Genetic characterization of *Toxoplasma gondii* from Brazilian wildlife revealed abundant new genotypes. Int. J. Parasitol. Parasites Wildl. *3*, 276–283.

Wallon, M., Liou, C., Garner, P., and Peyron, F. (1999). Congenital toxoplasmosis: systematic review of evidence of efficacy of treatment in pregnancy. BMJ *318*, 1511–1514.

BY-NC-ND

Wallon, M., Garweg, J.G., Abrahamowicz, M., Cornu, C., Vinault, S., Quantin, C., Bonithon-Kopp, C., Picot, S., Peyron, F., and Binquet, C. (2014). Ophthalmic outcomes of congenital toxoplasmosis followed until adolescence. Pediatrics *133*, e601-608.

Wang, Y., and Yin, H. (2014). Research progress on surface antigen 1 (SAG1) of *Toxoplasma gondii*. Parasit. Vectors *7*, 180.

Webb, S.D. (2006). The great american biotic interchange: patterns and processes. Ann. Mo. Bot. Gard. *93*, 245–257.

Wendte, J.M., Miller, M.A., Lambourn, D.M., Magargal, S.L., Jessup, D.A., and Grigg, M.E. (2010). Self-mating in the definitive host potentiates clonal outbreaks of the apicomplexan parasites Sarcocystis neurona and *Toxoplasma gondii*. PLoS Genet. *6*, e1001261.

Wendte, J.M., Gibson, A.K., and Grigg, M.E. (2011). Population genetics of *Toxoplasma gondii*: New perspectives from parasite genotypes in wildlife. Vet. Parasitol. *182*, 96–111.

White, N.J., Pukrittayakamee, S., Hien, T.T., Faiz, M.A., Mokuolu, O.A., and Dondorp, A.M. (2014). Malaria. Lancet Lond. Engl. *383*, 723–735.

Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human infectious diseases. Nature *447*, 279–283.

Yang, Z., and Rannala, B. (2012). Molecular phylogenetics: principles and practice. Nat. Rev. Genet. *13*, 303–314.

Yang, T., Deng, H.-W., and Niu, T. (2014). Critical assessment of coalescent simulators in modeling recombination hotspots in genomic sequences. BMC Bioinformatics *15*, 3.

BY-NC-ND

Annexes

Annexe 1. Article : Phylogeography of Toxoplasma gondii Points to a South American Origin 96

Annexe 2. Supplementary Information de l'article Annexe 1 109

Annexe 2.1. S1 Table : Year of isolation, geographic data, host origin, and microsatellite analysis of 168 *Toxoplasma gondii* isolates used in this study 109

Annexe 2.2. S2 Table : Polymorphism of different *Toxoplasma gondii* DNA regions screened in GenBank..... 131

Annexe 2.3. S3 Table : Polymorphism and composition of the 13 *Toxoplasma gondii* populations..... 134

Annexe 2.4. S4 Table : Position, size and primers of loci used in this study 135

Annexe 2.5. S5 Table : List of haplotypes..... 136

Annexe 2.6. S6 Table : Descriptive statistics of the data. 139

Annexe 2.7. S1 Figure 140

Annexe 2.8. S2 Figure 141

Annexe 2.9. S3 Figure 142

BY-NC-ND

Annexe 1. Article : Phylogeography of Toxoplasma gondii Points to a South American Origin

Article soumis à Infection, Genetics and Evolution (toujours en revue au 9 novembre 2016).

BY-NC-ND

Phylogeography of *Toxoplasma gondii* Points to a South American Origin

Emilie Bertranpetit[a], Thibaut Jombart[b], Emmanuel Paradis[c], Hilda Pena[d], Jitender Dubey[e], Chunlei Su[f], Aurélien Mercier[a], Sébastien Devillard[g,1, 2], Daniel Ajzenberg[h,1,2]

[a]*INSERM UMR_S 1094, Neuroépidémiologie Tropicale, Laboratoire de Parasitologie-Mycologie, Faculté de Médecine, Université de Limoges, Limoges, 87025, France*
[b]*MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, United Kingdom*
[c]*Institut des Sciences de l'Évolution, Université Montpellier/CNRS/IRD/EPHE, Place Eugène Bataillon – CC 065, 34095 Montpellier cédex 05, France*
[d]*Departamento de Medicina Veterinária Preventiva e Saúde Animal, Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, Brazil*
[e]*United States Department of Agriculture, Agricultural Research Service, Beltsville Agricultural Research Center, Animal Parasitic Diseases Laboratory, Building 1001, Beltsville, Maryland, 20705-2350, USA*
[f]*Department of Microbiology, University of Tennessee, Knoxville, Tennessee, 37996-0845, USA*
[g]*Université de Lyon, F-69000, Lyon ; Université Lyon 1 ; CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France*
[h]*Centre National de Référence (CNR) Toxoplasmose / Toxoplasma Biological Resource Center (BRC), Centre Hospitalier-Universitaire Dupuytren, Limoges, 87042, France and INSERM UMR_S 1094, Neuroépidémiologie Tropicale, Laboratoire de Parasitologie-Mycologie, Faculté de Médecine, Université de Limoges, Limoges, 87025, France*

[1]Correspondence and requests for materials should be addressed to S.D (email: sebastien.devillard@univ-lyon1.fr Tel: +33472448111) and D.A. (email: ajz@unilim.fr Tel: +33555056160)

[2] Authors contributed equally

BY-NC-ND

Abstract
Toxoplasma gondii, a protozoan found ubiquitously in mammals and birds, is the etiologic agent of toxoplasmosis, a disease causing substantial Public Health burden worldwide, including about 200,000 new cases of congenital toxoplasmosis each year. Clinical severity has been shown to vary across geographical regions, with South America exhibiting the highest burden. Unfortunately, the drivers of these heterogeneities are still poorly understood, and the geographical origin and historical spread of the pathogen worldwide are currently uncertain. A worldwide sample of 168 *T. gondii* isolates gathered in 13 populations was sequenced for five fragments of genes (140 single nucleotide polymorphisms from 3,153 bp per isolate). Phylogeny based on Maximum likelihood methods with estimation of the time to the most recent common ancestor (TMRCA) and geostatistical analyses were performed for inferring the putative origin of *T. gondii.* We show that extant strains of the pathogen likely evolved from a South American ancestor, around 1.5 million years ago, and reconstruct the subsequent spread of the pathogen worldwide. This emergence is much more recent than the appearance of ancestral *T. gondii*, believed to have taken place about 11 My ago, and follows the arrival of felids in this part of the world. We posit that an ancestral lineage of *T. gondii* likely arrived in South America with felids and that the evolution of oral infectivity through carnivorism and the radiation of felids in this region enabled a new strain to outcompete the ancestral lineage and undergo a pandemic radiation.

Keywords: *Toxoplasma gondii*, phylogeography, Maximum likelihood phylogeny, time to the most recent common ancestor (TMRCA), genetic diversity

1. Introduction

Toxoplasmosis imposes a substantial disease burden across the world. Serological studies demonstrate its presence in virtually every country, with seroprevalence exceeding 60% in some parts of South America, Africa, and South-East Asia (Pappas et al., 2009). While asymptomatic in most patients, toxoplasmosis is a major cause of uveitis in immunocompetent patients and a potentially life-threatening illness in immunocompromised patients and fetuses (Montoya and Liesenfeld, 2004). Congenital toxoplasmosis alone has a yearly global incidence of about 200,000 cases, causing a burden exceeding 1 million disability-adjusted life year (DALYs) (Torgerson and Mastroiacovo, 2013). The overall disease burden attributable to all forms of toxoplasmosis is certainly far greater, and highest in South America where ocular toxoplasmosis is unusually frequent and severe (Glasner et al., 1992; de-la-Torre et al., 2008; Gilbert et al., 2008; Torgerson and Mastroiacovo, 2013).

The extent to which host parasite genetics, host immune status, and exposure rate contribute to the increased severity of toxoplasmosis in South America is unclear, but differences in the genetic makeup of *T. gondii* strains are likely to play a major role (Khan et al., 2006; Gilbert et al., 2008). In North America, Europe, Africa, and Asia, the population structure of *T. gondii* is dominated by a few prevalent clonal strains, whereas much greater genetic diversity is seen in tropical South America where the populations lack sign of recent genetic bottleneck and clonal structure seen in the other parts of the world (Shwab et al., 2014; Lorenzi et al., 2016). Unfortunately, the drivers of the pathogen's genetic diversity are still poorly understood, and the origin of extant lineages of *T. gondii* remains controversial. Recent work suggested a potential South American origin, while another study advocated that co-migration with felids led to the divergence of South American strains from pre-existing North American ones (Lehmann et al., 2006; Khan et al., 2007). The estimation of the time to the most recent common

BY-NC-ND

ancestor (TMRCA) of extant lineages is also disputed, with estimates ranging from 150,000 to 10^7 years (Morrison, 2005; Khan et al., 2007).
In the present study, using a large collection of genetic sequences of *T. gondii* sampled worldwide, we reconstructed the phylogeography of *T. gondii* as a basis to address the controversial questions regarding the evolution of this parasite and its geographical origin.

2. Materials and Methods

2.1. Collection of *T. gondii* strains and selection of markers

A total of 168 *T. gondii* strains from 13 populations collected worldwide in North America, South America, the Caribbean, Europe, Asia, and Africa, were used in this study (Supplementary information and S1 Table). Our collection was specifically designed to ensure extensive geographic coverage of strains clustered in true populations, and includes a large number of strains from Africa and China, which were so far underrepresented or absent in the previous studies (Lehmann et al., 2006; Khan et al., 2007). A preliminary genetic analysis with 15 microsatellite markers (Ajzenberg et al., 2010) was performed to exclude clones of strains sampled in the same area. Each isolate was sequenced in both directions for five markers (*GRA6, GRA7*, *SAG3*, *UPRT1* and *UPRT7*) that had the highest polymorphic rates after a preliminary analysis of genetic polymorphism of 30 fragments of genes retrieved from GenBank and ToxoDB (Supplementary information and S2 Table).

2.2. Phylogenetic analysis.

Hammondia hammondi is the most closely parasite related to *T. gondii* and was used as an outgroup in phylogenetic analyses. Sequences of the *H. hammondi* strain H.H.34 corresponding to *T. gondii GRA6*, *GRA7*, *SAG3*, *UPRT1* and *UPRT7* sequences were retrieved from GenBank and ToxoDB, and aligned with MUSCLE (Edgar, 2004). We used *ape* and *pegas* R packages to extract haplotypes and build phylogenetic trees with three distance-based methods: NJ, BIONJ, and FastME which were used as starting trees for the ML analyses. Maximum likelihood phylogenetic analyses were performed with the R package *phangorn* using four partitions of the sequence data by crossing two criteria: exons *vs.* introns on one side, and *GRA6*, *GRA7* and *SAG3 vs. UPRT1* and *UPRT7* on the other (Supplementary information). A GTR + Γ + I model was used with parameters that could vary among data partitions. The different model fits performed with *phangorn* were compared with AIC. The three trees obtained with the distance-based methods were all tested as initial trees.

2.3. Geostatistical analyses. We used the same geostatistics approach for inferring the putative origin of *T. gondii* as previously used for uncovering the origins of *P. falciparum* (Tanabe et al., 2010, 2013a, 2013b; Mita and Jombart, 2015). We implemented this approach in the R package *geoGraph* (http://thibautjombart/geograph), in which we provided extensive documentation replicating the analyses described below using publicly available data (Cann et al., 2002). The method implemented in *geoGraph* relies on the idea that migration events result in successive bottlenecks which reduce the genetic diversity within populations as they are located further away from the origin (Tanabe et al., 2010). Accordingly, we expect to observe a negative correlation between within-population diversity and the distance from the origin. While in practice the true origin is often unknown, one can infer the most plausible origin by assessing this relationship for a number of candidate origins, and retaining the origin yielding the strongest negative correlation. This method requires two types of distances, genetic and spatial, to be computed. Here, the genetic diversity was mostly structured by

BY-NC-ND

varying frequencies of a small number of haplotypes within populations (S3 Table and Fig 1). Therefore, we used haplotype richness (i.e. number of distinct haplotypes) as a measure of diversity within populations. Spatial distances through landmasses were computed using *geoGraph*. The package models movements on the surface of the Earth using a spherical, pseudo-regular grid with approximately 40,000 nodes. Each node possesses an 'habitat' attribute, here used to distinguish landmasses from seas. Shortest path between locations were computed using the dijkstra algorithm (Jungnickel, 2013) implemented in the R package RBGL (Edmonds et al., 2006; Carey et al., 2011). To define candidate origins, 1,800 combinations of regularly spaced longitudes and latitudes were used to cover the globe, which resulted in 433 non-redundant locations on landmasses on the grid used by *geoGraph*. For each location, the shortest path through each sampled population was identified, and the corresponding distance computed in kilometers. These distances were then used to assess patterns of decrease of genetic diversity from the putative origin using simple linear regression. The most likely origin was inferred as the location which yielded the most negative correlation between geographic distances and haplotype diversity within populations.

2.4. Time to the most recent common ancestor (TMRCA). In order to estimate the time to the most recent common ancestor (TMRCA) of *T. gondii*, we used two different approaches: a simple molecular dating method based on the divergence with *H. hammondi* which is estimated to be around 11 My, and a coalescent approach using the expectation of TMRCA which is equal to twice the effective population size (Ne). Both approaches need an estimate of the mutation rate (μ), and the second one also needs an estimate of the population parameter θ (= 2 μ Ne). We estimated μ for the non-coding introns of *UPRT1* and *UPRT7*. We did two neutrality tests: D's Tajima and the R2 test. We calculated a standard-error of this estimate using the variance of ratio of two random variables, here the number of mutations per site and twice the time of divergence. The former was estimated with a Tamura-Nei distance with its associated variance, and the latter was 11 My with an arbitrary sd = 1 My. The population parameter θ was estimated in two ways: with a Markov chain Monte Carlo (MCMC) approach as implemented in the R package *coalescentMCMC*, and with the nucleotide diversity (π) calculated with *pegas*. Both ways calculate the standard-error of the estimate of θ. In the end, three estimates of TMRCA were obtained with their respective 95% confidence interval (CI).

3. Results

3.1. Genetic diversity.

Sequences of the five markers represented a total of 3,153 bp per isolate, including 140 variable sites. Without taking into account sites with gaps, 26, 30, 27, 32, and 25 SNPs were identified in the *GRA6* (607 bp), *GRA7* (677 bp), *SAG3* (638 bp), *UPRT1* (574 bp), and *UPRT7* (657 bp) genes, respectively (S4 Table and S1 Fig). Strains from the Caribbean, Guiana shield, Northeast Brazil, and Southeast Brazil showed more sequence polymorphism with 65, 90, 59, and 65 SNPs, respectively (S3 Table). Polymorphism was lower in the African and European populations with a number of SNPs ranging from one to 39. Polymorphism was intermediate in the Asian and North American populations with 41, 60, 47, and 51 SNPs in Turkey, China, Minnesota, and Pennsylvania, respectively.

Data concatenation revealed 60 haplotypes. The number of haplotypes ranged from four in Europe to 32 in South America. Of the 32 genotypes in the 44 strains from South America, 29 were endemic in South America whereas only three were common in other populations (one in the Caribbean, one in Africa, and one both in Asia and Africa). In contrast, of the four genotypes in the 34 strains from Europe, three were common in other populations (one in Asia

BY-NC-ND

and Africa, one in North America, Asia and Africa, and one in the Caribbean, North America, Asia and Africa) and the unique genotype differed by only one SNP from the one common to North America, Asia, and Africa (S5 Table). The higher values of haplotype diversity, estimated from concatenation of the five markers by the number of haplotypes divided by the number of isolates, were observed in South America and the Caribbean, whereas the lower values were observed in Europe, Asia, and Africa (S3 Table). Haplotype diversity was intermediate in North America. Overall the highest genetic diversity was found in South-America.

3.2. Phylogeny of *T. gondii* strains

Preliminary analyses of genetic diversity (Supplementary information, S6 Table and S2 and S3 Figs.) revealed different mutation patterns and rates of evolution between coding and non-coding segments, and for the two *UPRT* genes compared to the others. Accordingly, we defined four partitions of the sequence data crossing these two categories, and reconstructed separate phylogenies by maximum likelihood (ML) to investigate potential phylogenetic incongruence (Som, 2015). Statistical tests and examination of model selection criterion (AIC) revealed the existence of distinct topologies (Fig. 1), suggesting that these sequence partitions have undergone different evolutionary histories and selective pressures. Interestingly, only South American strains were consistently placed at a basal position (close to the root) in all topologies. To investigate this pattern further and identify the common evolutionary history of these genes, a consensus topology was inferred from the four ML-partitioned topologies (Fig. 2). This new tree supported the more ancestral status of South American isolates, with 35 out of 44 samples located at the root of the tree. However, as expected in the presence of conflicting phylogenetic signal, this tree was only partially resolved, and strains from other locations (China: 12 samples; Africa: 7 samples) also belonged to the large basal multifurcation.

BY-NC-ND

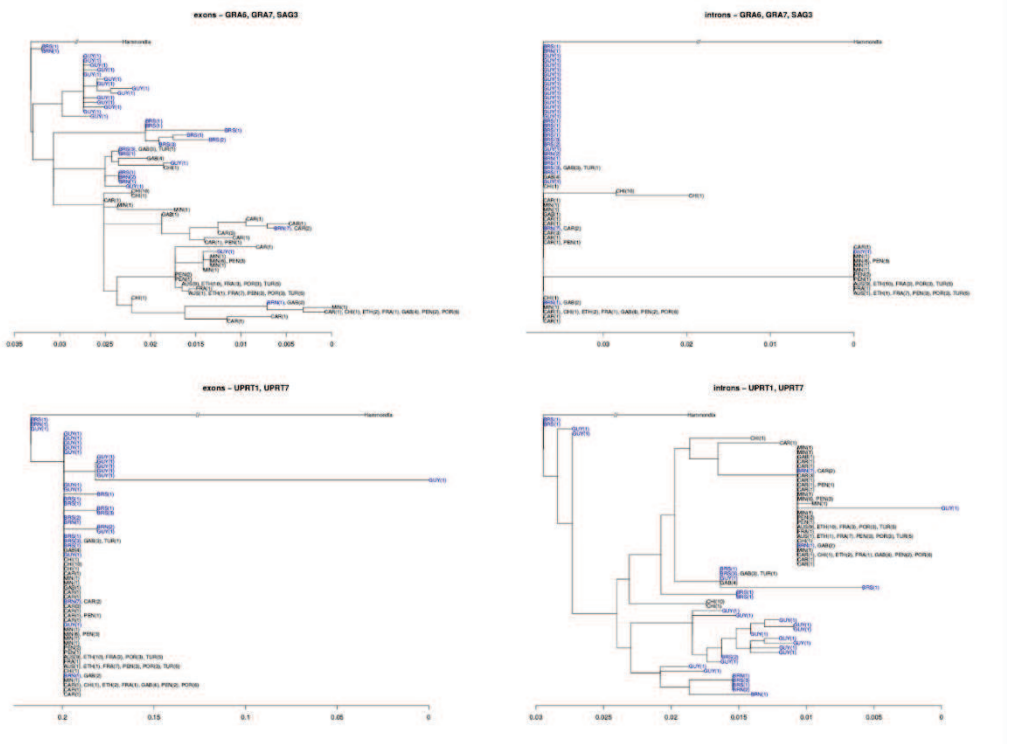


Fig 1. Maximum likelihood phylogenies for 5 molecular markers.
Different topologies are supported by the four data partitions. The different rates of evolution are indicated by the different scales of branch lengths. The trees were rooted with the outgroup and the long branch leading to the outgroup is shown broken. The strains from South America are shown in blue.

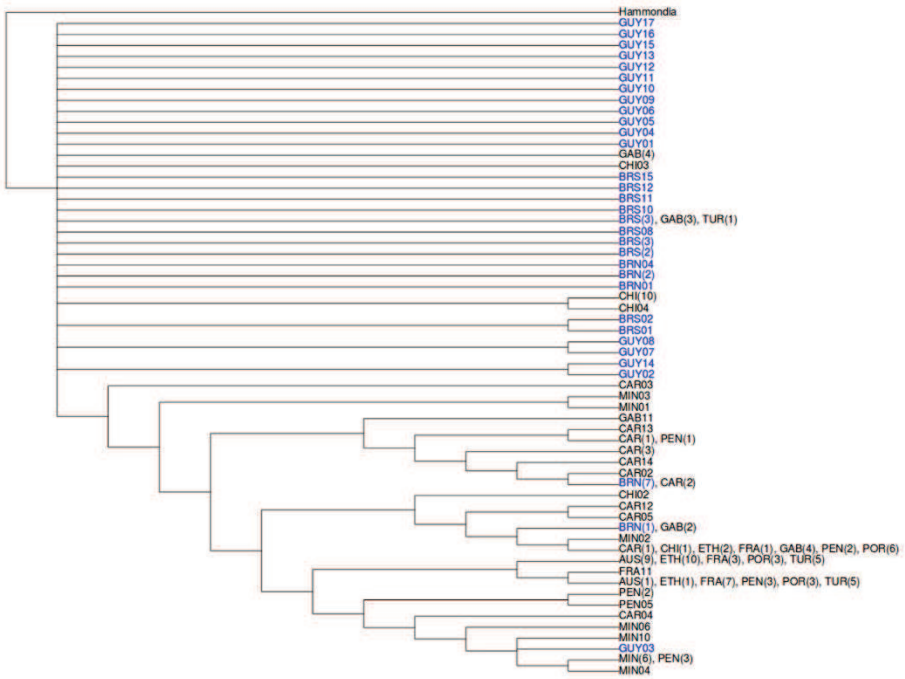


Fig. 2. Consensus tree built from the four trees in Figure 1.
The strains from South America are shown in blue. If a haplotype (unique sequence over the five genes) was observed in a single strain, its reference number is shown; otherwise the geographic origins of the strains (and frequencies within parentheses) are indicated.

BY-NC-ND

3.3. On the geographic origin of *T. gondii*

As a complementary analysis, we used a geostatistical approach previously employed for identifying the origin of *Plasmodium falciparum*, the main etiologic agent of malaria (Tanabe et al., 2010, 2013b; Mita and Jombart, 2015). This method identifies likely geographic origins as the locations from which patterns of decrease in genetic diversity, expected to be observed due to repeated migration and founder effects, are most consistent (Tanabe et al., 2010). Because of the low level of polymorphism observed in the sequenced genes and the highly clonal nature of *T. gondii*, haplotype richness was used as a measure of genetic diversity within populations (S3 Table). Testing a large number of hypothetical origins across the world, this approach identified South America, and more specifically Colombia as the most likely origin (r=-0.81, p=0.9x10^{-4}, Fig. 3). While substantial uncertainty remains about the exact location, this analysis brings strong support to a South American origin for *T. gondii* suggested by the phylogenetic approach. Our results further suggest that *T. gondii* initially spread through the Americas and then colonized Asia and Europe via the Bering Strait, before entering Africa through two different migration routes (Fig. 3).

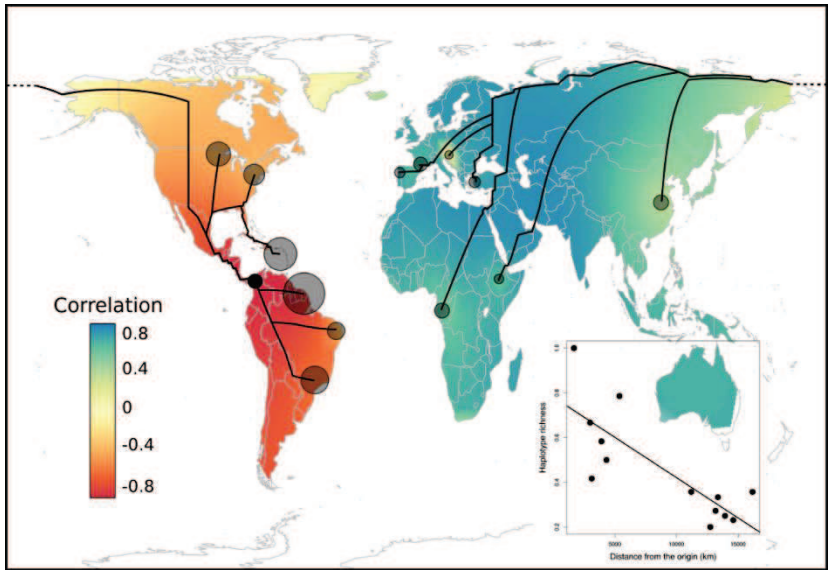


Fig 3. Historical spread of *Toxoplasma gondii* inferred from five molecular markers. Haplotype diversity within populations (grey circles) is correlated to the geographic distance from various origins, indicated using color shading (see legend). Strong negative correlations (red shading) indicate substantial loss of diversity inferred from the corresponding location. The best supported origin (plain dot) corresponds to strongest loss of diversity (i.e. lowest correlation), indicated in inset. The corresponding shortest path from the origin to the various populations is indicated by plain black lines.

3.4. Time to the most recent common ancestor (T_{MRCA}) of *T. gondii*.

To understand the processes which may have led to a pandemic radiation of *T. gondii* from South America, the emergence of this ancestral, highly successful lineage has to be dated. To this end, we derived estimates of time to the most recent common ancestor (MRCA) of the extant strains using standard molecular approaches. While confidence intervals indicated substantial uncertainty, overall results suggest that the MRCA of *T. gondii* appeared around 1.5 My ago (Table 1). This emergence is much more recent that the existence of *T. gondii*

BY-NC-ND

itself, estimated to have diverged from its closest ancestor *Hammondia hammondi* some 11 My ago (Morrison et al., 2004).

Table 1. Estimates of the time to the most recent common ancestor (TMRCA) of *Toxoplasma gondii* with three different methods on the introns of *UPRT* genes.

Method	TMRCA (Ma)	[95% CI]
Molecular dating	1.59	[0.00–3.46]
Coalescent	1.26	[0.94–1.57]
Nuc. div. (π)	1.20	[0.00–2.44]

Nuc. div.: nucleotide diversity.

4. Discussion

A major event occurred in the evolutionary history of *T. gondii* which led to a selective sweep about 1.5 My ago. We hypothesized that an ancestral form of *T.gondii* was introduced in South America through the migration of Felidae after the emergence of the Isthmus of Panama about 2 to 3 My ago, at the end of Pliocene (O'Brien et al., 2008). It is believed that Felidae species quickly expanded after their arrival and diversified into the "ocelot" lineage in South America. Interestingly the Muridae, potential intermediate hosts for *T. gondii*, also showed extensive diversification in South America with the appearance of several genera (Webb, 2006) around the same time. As previously suggested (Webb, 2006), we posit that this expansion, diversification and mixing of host populations favoured the accumulation of genetic diversity in *T. gondii*, which eventually led to a selective sweep by a highly successful mutant lineage.

The selective pressures underlying this selective sweep can be debated. Previous work attributed this radiation to the emergence of transmission through carnivorism (i.e., oral infectivity of tissue cysts) between intermediate hosts in clonal strains 10,000 years ago (Su et al., 2003). However, oral infectivity was shown to be also a trait of many South American strains (Carme et al., 2002; Khan et al., 2007). Because South American strains were the first to diverge from the MRCA, it is likely that transmission by carnivorism evolved earlier than the apparition of clonal lineages. This trait conferred a better transmission of the current form of *T. gondii* which likely outcompeted the ancestral form arriving in South America with Felidae. The transmission of *T. gondii* between its different hosts would allow some genotypes to migrate to North America, then to go through the Bering Strait to colonize Asia, Europe and Africa. The current population structure of *T. gondii* with a predominance of a few successful clonal strains in Africa, Asia and Europe, is likely to be the consequence of the recent expansion of the domestic cat, an Old World species until the sixteenth century, that tremendously amplified a specific subset of pre-adapted genotypes (Müller and Howard, 2016).

In addition to being the likely origin of modern *T. gondii* strains, South America also suffers from the highest burden of toxoplasmosis. Prevalence, incidence, and severity of acquired and congenital ocular toxoplasmosis (OT) in some areas of Brazil, Colombia, and Argentina are

BY-NC-ND

considerably higher than anywhere else, which makes OT a genuine public health issue in South America (Glasner et al., 1992; de-la-Torre et al., 2008; Gilbert et al., 2008; Rudzinski et al., 2016). Because South America is also the hotspot of *T. gondii*'s genetic diversity, it has been hypothesized that severe forms of toxoplasmosis may be the consequence of poor adaptation of the human host to the unusual diversity of strains in this part of the world, resulting in impaired immune response and, thus, a more aggressive disease (Khan et al., 2006; Gilbert et al., 2008; Demar et al., 2012; de-la-Torre et al., 2013; Rudzinski et al., 2016). The societal and economic costs of care for symptomatic cases of congenital toxoplasmosis can be considerable but the cost-effectiveness of national routine prenatal screening and treatment program are still debated (Wallon et al., 1999; Jones et al., 2014). There is a need for randomized placebo-controlled trials to help determine the effectiveness of these interventions.

5. Conclusion

Our reconstruction of *T. gondii*'s phylogeography provides a new framework for understanding patterns of genetic diversity in sampled populations of the parasite, and for predicting diversity in unsampled locations. Because genetic diversity seems to impact directly the severity of the disease, our results can be used as a basis for explaining geographic heterogeneities in disease burden, and identifying priority targets for potential future interventions.

ACKNOWLEDGEMENTS

We are thankful to github (http://github.com/) and travis (http://travis-ci.org/) for providing great resources for the development of the R package *geoGraph*. We are indebted to Endrias Zewdu Gebremedhin, Hüseyin Can, Mert Döşkaya, Yüksel Gürüz, Susana Sousa, Lénaïg Halos, Min Li, François Peyron, and the members of the French National Reference Center for toxoplasmosis (Nicole Desbois, Saadia Azi, Hélène Yera, Magalie Demar, Bernard Carme, Stéphane Simon, Denis Blanchet, Rachida Boukhari, Isabelle Villena, Dominique Aubert) for their contribution to the collection of 168 strains that were included in the analysis.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

The sequences reported in this study have been deposited in the GenBank database (accession n° KU598987-KU599154 for *GRA6*, accession n°KU599155-KU599322 for *GRA7*, accession n°KU599323-KU599490 for *SAG3*, accession n°KU599491-KU599658 for *UPRT1* and accession n°KU599659-KU599826 for *UPRT7*).

BY-NC-ND

References

Ajzenberg, D., Daniel, A., Frédéric, C., Aurélien, M., Philippe, V., Marie-Laure, D., 2010. Genotyping of *Toxoplasma gondii* Isolates with 15 Microsatellite Markers in a Single Multiplex PCR Assay. J. Clin. Microbiol. 48, 4641–4645.

Cann, H.M., de Toma, C., Cazes, L., Legrand, M.-F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., Chen, Z., Chu, J., Carcassi, C., Contu, L., Du, R., Excoffier, L., Ferrara, G.B., Friedlaender, J.S., Groot, H., Gurwitz, D., Jenkins, T., Herrera, R.J., Huang, X., Kidd, J., Kidd, K.K., Langaney, A., Lin, A.A., Mehdi, S.Q., Parham, P., Piazza, A., Pistillo, M.P., Qian, Y., Shu, Q., Xu, J., Zhu, S., Weber, J.L., Greely, H.T., Feldman, M.W., Thomas, G., Dausset, J., Cavalli-Sforza, L.L., 2002. A human genome diversity cell line panel. Science 296, 261–262.

Carey, V., Long, L., Gentleman, R., 2011. RBGL: An interface to the BOOST graph library. R package version 1, 124.

Carme, B., Bissuel, F., Ajzenberg, D., Bouyne, R., Aznar, C., Demar, M., Bichat, S., Louvel, D., Bourbigot, A.M., Peneau, C., Neron, P., Darde, M.L., 2002. Severe Acquired Toxoplasmosis in Immunocompetent Adult Patients in French Guiana. J. Clin. Microbiol. 40, 4037–4044.

de-la-Torre, A., López-Castillo, C.A., Gómez-Marín, J.E., 2008. Incidence and clinical characteristics in a Colombian cohort of ocular toxoplasmosis. Eye 23, 1090–1093.

de-la-Torre, A., Sauer, A., Pfaff, A.W., Bourcier, T., Brunet, J., Speeg-Schatz, C., Ballonzoli, L., Villard, O., Ajzenberg, D., Sundar, N., Grigg, M.E., Gomez-Marin, J.E., Candolfi, E., 2013. Severe South American ocular toxoplasmosis is associated with decreased Ifn-γ/Il-17a and increased Il-6/Il-13 intraocular levels. PLoS Negl. Trop. Dis. 7, e2541.

Demar, M., Hommel, D., Djossou, F., Peneau, C., Boukhari, R., Louvel, D., A.-M., B., Nasser, V., Ajzenberg, D., M.-L., D., Carme, B., 2012. Acute toxoplasmoses in immunocompetent patients hospitalized in an intensive care unit in French Guiana. Clin. Microbiol. Infect. 18, E221–E231.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Edmonds, N., Breuer, A., Gregor, D., Lumsdaine, A., 2006. Single-source shortest paths with the parallel boost graph library. The Ninth DIMACS Implementation Challenge: The Shortest Path Problem, Piscataway, NJ 219–248.

Gilbert, R.E., Freeman, K., Lago, E.G., Bahia-Oliveira, L.M.G., Tan, H.K., Wallon, M., Buffolano, W., Stanford, M.R., Petersen, E., European Multicentre Study on Congenital Toxoplasmosis (EMSCOT), 2008. Ocular sequelae of congenital toxoplasmosis in Brazil compared with Europe. PLoS Negl. Trop. Dis. 2, e277.

BY-NC-ND

Glasner, P.D., Silveira, C., Kruszon-Moran, D., Martins, M.C., Burnier Júnior, M., Silveira, S., Camargo, M.E., Nussenblatt, R.B., Kaslow, R.A., Belfort Júnior, R., 1992. An unusually high prevalence of ocular toxoplasmosis in southern Brazil. Am. J. Ophthalmol. 114, 136–144.

Jones, J.L., Parise, M.E., Fiore, A.E., 2014. Neglected parasitic infections in the United States: toxoplasmosis. Am. J. Trop. Med. Hyg. 90, 794–799.

Jungnickel, D., 2013. Graphs, Networks and Algorithms:, Algorithms and Computation in Mathematics . Springer Berlin Heidelberg.

Khan, A., Fux, B., Su, C., Dubey, J.P., Darde, M.L., Ajioka, J.W., Rosenthal, B.M., Sibley, L.D., 2007. Recent transcontinental sweep of *Toxoplasma gondii* driven by a single monomorphic chromosome. Proc. Natl. Acad. Sci. U. S. A. 104, 14872–14877.

Khan, A., Jordan, C., Muccioli, C., Vallochi, A.L., Rizzo, L.V., Belfort, R., Jr, Vitor, R.W.A., Silveira, C., Sibley, L.D., 2006. Genetic divergence of *Toxoplasma gondii* strains associated with ocular toxoplasmosis, Brazil. Emerg. Infect. Dis. 12, 942–949.

Lehmann, T., Marcet, P.L., Graham, D.H., Dahl, E.R., Dubey, J.P., 2006. Globalization and the population structure of *Toxoplasma gondii*. Proc. Natl. Acad. Sci. U. S. A. 103, 11423–11428.

Lorenzi, H., Khan, A., Behnke, M.S., Namasivayam, S., Swapna, L.S., Hadjithomas, M., Karamycheva, S., Pinney, D., Brunk, B.P., Ajioka, J.W., Ajzenberg, D., Boothroyd, J.C., Boyle, J.P., Dardé, M.L., Diaz-Miranda, M.A., Dubey, J.P., Fritz, H.M., Gennari, S.M., Gregory, B.D., Kim, K., Saeij, J.P.J., Su, C., White, M.W., Zhu, X.-Q., Howe, D.K., Rosenthal, B.M., Grigg, M.E., Parkinson, J., Liu, L., Kissinger, J.C., Roos, D.S., Sibley, L.D., 2016. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. Nat. Commun. 7, 10147.

Mita, T., Jombart, T., 2015. Patterns and dynamics of genetic diversity in *Plasmodium falciparum*: what past human migrations tell us about malaria. Parasitol. Int. 64, 238–243.

Montoya, J.G., Liesenfeld, O., 2004. Toxoplasmosis. Lancet 363, 1965–1976.

Morrison, D.A., 2005. How old are the extant lineages of *Toxoplasma gondii*? Parassitologia 47, 205–214.

Morrison, D.A., Set, B., Per, T., Ulrich, W., Jörg, K., Mattsson, J.G., 2004. The current status of the small subunit rRNA phylogeny of the coccidia (Sporozoa). Int. J. Parasitol. 34, 501–514.

Müller, U.B., Howard, J.C., 2016. The impact of *Toxoplasma gondii* on the mammalian genome. Curr. Opin. Microbiol. 32, 19–25.

O'Brien, S.J., Johnson, W., Driscoll, C., Pontius, J., Pecon-Slattery, J., Menotti-Raymond, M., 2008. State of cat genomics. Trends Genet. 24, 268–279.

BY-NC-ND

Pappas, G., Roussos, N., Falagas, M.E., 2009. Toxoplasmosis snapshots: global status of *Toxoplasma gondii* seroprevalence and implications for pregnancy and congenital toxoplasmosis. Int. J. Parasitol. 39, 1385–1394.

Rudzinski, M., Khoury, M., Couto, C., Ajzenberg, D., 2016. Reactivation of Ocular Toxoplasmosis in Non-Hispanic Persons, Misiones Province, Argentina. Emerg. Infect. Dis. 22, 912–913.

Shwab, E.K., Zhu, X.-Q., Majumdar, D., Pena, H.F.J., Gennari, S.M., Dubey, J.P., Su, C., 2014. Geographical patterns of *Toxoplasma gondii* genetic diversity revealed by multilocus PCR-RFLP genotyping. Parasitology 141, 453–461.

Som, A., 2015. Causes, consequences and solutions of phylogenetic incongruence. Brief. Bioinform. 16, 536–548.

Su, C., Evans, D., Cole, R.H., Kissinger, J.C., Ajioka, J.W., Sibley, L.D., 2003. Recent Expansion of *Toxoplasma* Through Enhanced Oral Transmission. Science 299, 414–416.

Tanabe, K., Jombart, T., Horibe, S., Palacpac, N.M.Q., Honma, H., Tachibana, S.-I., Nakamura, M., Horii, T., Kishino, H., Mita, T., 2013a. *Plasmodium falciparum* mitochondrial genetic diversity exhibits isolation-by-distance patterns supporting a sub-Saharan African origin. Mitochondrion 13, 630–636.

Tanabe, K., Mita, T., Jombart, T., Eriksson, A., Horibe, S., Palacpac, N., Ranford-Cartwright, L., Sawai, H., Sakihama, N., Ohmae, H., Nakamura, M., Ferreira, M.U., Escalante, A.A., Prugnolle, F., Björkman, A., Färnert, A., Kaneko, A., Horii, T., Manica, A., Kishino, H., Balloux, F., 2010. *Plasmodium falciparum* accompanied the human expansion out of Africa. Curr. Biol. 20, 1283–1289.

Tanabe, K., Mita, T., Palacpac, N.M.Q., Arisue, N., Tougan, T., Kawai, S., Jombart, T., Kobayashi, F., Horii, T., 2013b. Within-population genetic diversity of *Plasmodium falciparum* vaccine candidate antigens reveals geographic distance from a Central sub-Saharan African origin. Vaccine 31, 1334–1339.

Torgerson, P.R., Mastroiacovo, P., 2013. The global burden of congenital toxoplasmosis: a systematic review. Bull. World Health Organ 91, 501–508..

Wallon, M., Liou, C., Garner, P., Peyron, F., 1999. Congenital toxoplasmosis: systematic review of evidence of efficacy of treatment in pregnancy. BMJ 318, 1511–1514.

Webb, S.D., 2006. The Great American Biotic Interchange: patterns and processes 1. Ann. Mo. Bot. Gard. 93, 245–257.

BY-NC-ND

Annexe 2. Supplementary Information de l'article Annexe 1

Annexe 2.1. S1 Table : Year of isolation, geographic data, host origin, and microsatellite analysis of 168 *Toxoplasma gondii* isolates used in this study

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgSpEt23	ETH01	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Sheep (Ovis aries)
TgGtEt05	ETH02	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Goat (Capra aegagrus hircus)
TgSpEt18	ETH03	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Sheep (Ovis aries)
TgSpEt25	ETH04	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Sheep (Ovis aries)
TgGtEt08	ETH05	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Goat (Capra aegagrus hircus)
TgSpEt06	ETH06	2011	Africa / Ethiopia	Ethiopia	8.953888	37.859722	Sheep (Ovis aries)
TgSpEt08	ETH07	2011	Africa / Ethiopia	Ethiopia	8.953888	37.859722	Sheep (Ovis aries)
TgGtEt19	ETH08	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Goat (Capra aegagrus hircus)
TgGtEt01	ETH09	2011	Africa / Ethiopia	Ethiopia	8.953888	37.859722	Goat (Capra aegagrus hircus)
TgGtEt14	ETH10	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Goat (Capra aegagrus hircus)
TgGtEt02	ETH11	2011	Africa / Ethiopia	Ethiopia	8.953888	37.859722	Goat (Capra aegagrus hircus)
TgGtEt04	ETH12	2011	Africa / Ethiopia	Ethiopia	8.953888	37.859722	Goat (Capra aegagrus hircus)
TgSpEt26	ETH13	2011	Africa / Ethiopia	Ethiopia	8.98	39.927499	Sheep (Ovis aries)
GAB1-GAL-DOM06-(TgA105015)	GAB01	2007	Africa / Gabon	Gabon	-1.87445	12,7317	Chicken (Gallus gallus domesticus)
GAB1-GAL-DOM11-(TgA105016)	GAB02	2007	Africa / Gabon	Gabon	-1.87668	12,73167	Chicken (Gallus gallus domesticus)
GAB1-CAP-AEG06-(TgA105011)	GAB03	2007	Africa / Gabon	Gabon	-1.87221	12,73042	Goat (Capra aegagrus hircus)
GAB1-GAL-DOM13-(TgA105018)	GAB04	2007	Africa / Gabon	Gabon	-1.88041	12,73082	Chicken (Gallus gallus domesticus)
GAB2-GAL-DOM06-(TgA105040)	GAB05	2007	Africa / Gabon	Gabon	0.60260	12,93903	Chicken (Gallus gallus domesticus)
GAB4-GAL-DOM01-(TgA105053)	GAB06	2007	Africa / Gabon	Gabon	-0.10031	11,60844	Chicken (Gallus gallus domesticus)
GAB2-GAL-DOM02-(TgA105004)	GAB07	2007	Africa / Gabon	Gabon	0.57989	12,88778	Chicken (Gallus gallus domesticus)
GAB2-GAL-DOM01-(TgA105037)	GAB08	2007	Africa / Gabon	Gabon	0.57989	12,88778	Chicken (Gallus gallus domesticus)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
GAB1-GAL-DOM10-(TgA105003)	GAB09	2007	Africa / Gabon	Gabon	-1.88393	12,73	Chicken (Gallus gallus domesticus)
GAB3-GAL-DOM08-(TgA105052)	GAB10	2007	Africa / Gabon	Gabon	0.42664	9,47009	Chicken (Gallus gallus domesticus)
GAB5-GAL-DOM03-(TgA105055)	GAB11	2007	Africa / Gabon	Gabon	-1.63918	13,58253	Chicken (Gallus gallus domesticus)
GAB3-GAL-DOM02-(TgA105001)	GAB12	2007	Africa / Gabon	Gabon	0.36838	9,48341	Chicken (Gallus gallus domesticus)
GAB3-GAL-DOM11-(TgA105043)	GAB13	2007	Africa / Gabon	Gabon	0.41103	9,43892	Chicken (Gallus gallus domesticus)
GAB5-GAL-DOM01-(TgA105002)	GAB14	2007	Africa / Gabon	Gabon	-1.64170	13,57267	Chicken (Gallus gallus domesticus)
TONT	CAR01	1992	Caribbean	Guadeloupe	16.26166	-61.512777	Human / Congenital
TgCatPr09	CAR02	2004	Caribbean	Puerto Rico	18.08139	-67.89444	Cat (Felis silvestris catus)
TgH16001	CAR03	2007	Caribbean	Martinique	14.6222	-61.0561	Human / AIDS
TgH40001	CAR04	2009	Caribbean	Haïti	18.5508	-72.2525	Human / AIDS
TgCatPr07	CAR05	2004	Caribbean	Puerto Rico	18.08139	-67.89444	Cat (Felis silvestris catus)
TgCatPr08	CAR06	2004	Caribbean	Puerto Rico	18.08139	-67.89444	Cat (Felis silvestris catus)
ENVL-2002-MAC	CAR07	2002	Caribbean	Barbados	13.15055	-59.5461	Macaque (Cynomolgus sp)
TgH40002	CAR08	2010	Caribbean	Guadeloupe	16.26166	-61.512777	Human / Congenital
TgH16010	CAR09	2013	Caribbean	Martinique	14.5761	-60.97333	Human / Congenital
TgCatStk07a	CAR10	2006	Caribbean	Saint Kitts and Nevis	17.29889	-62.72278	Cat (Felis silvestris catus)
TgH13006	CAR11	2005	Caribbean	Guadeloupe	16.26166	-61.512777	Human / BMT
TgCatPr06	CAR12	2004	Caribbean	Puerto Rico	18.08139	-67.89444	Cat (Felis silvestris catus)
TgCatPr10	CAR13	2004	Caribbean	Puerto Rico	18.08139	-67.89444	Cat (Felis silvestris catus)
TgH16012	CAR14	2014	Caribbean	Martinique	14.6222	-61.0561	Human / AIDS
TgWolfMN20	MIN01	2011	North America / Minnesota	USA	46,81237	-94,28989	Gray wolf (Canis lupus)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgWtdUs10	MIN02	2007	North America / Minnesota	USA	48.54694	-95.60250	White-tailed deer (Odocoileus virginianus)
TgWtdUs08	MIN03	2007	North America / Minnesota	USA	48.54694	-95.60250	White-tailed deer (Odocoileus virginianus)
TgWolfMN11	MIN04	2010	North America / Minnesota	USA	47,8083	-94,67914	Gray wolf (Canis lupus)
TgWolfMN29	MIN05	2011	North America / Minnesota	USA	46,58417	-92,51895	Gray wolf (Canis lupus)
TgWolfMN19	MIN06	2010	North America / Minnesota	USA	47,04742	-91,67799	Gray wolf (Canis lupus)
TgWolfMN13	MIN07	2010	North America / Minnesota	USA	45,96829	-92,76389	Gray wolf (Canis lupus)
TgWolfMN28	MIN08	2011	North America / Minnesota	USA	46,68661	-92,81987	Gray wolf (Canis lupus)
TgWolfMN26	MIN09	2011	North America / Minnesota	USA	46,66622	-94,77565	Gray wolf (Canis lupus)
TgWolfMN12	MIN10	2010	North America / Minnesota	USA	47,8763	-93,64966	Gray wolf (Canis lupus)
TgWolfMN27	MIN11	2011	North America / Minnesota	USA	46,70925	-94,73335	Gray wolf (Canis lupus)
TgWolfMN25	MIN12	2011	North America / Minnesota	USA	46,00503	-92,54156	Gray wolf (Canis lupus)
TgFoxPa03	PEN01	2011	North America / Pennsylvania	USA	39.8697222	-77.222	Red fox (Vulpes vulpes)
TgCoPa08	PEN02	2011	North America / Pennsylvania	USA	41.350277	-77.022	Coyote (Canis latrans)
TgCoPa05	PEN03	2011	North America / Pennsylvania	USA	40.9666	-78.435	Coyote (Canis latrans)
TgCoPa01	PEN04	2012	North America / Pennsylvania	USA	41.81472	-75.80388	Coyote (Canis latrans)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgFoxPa10	PEN05	2011	North America / Pennsylvania	USA	40.043888	-75.19666	Red fox (Vulpes vulpes)
TgCoPa07	PEN06	2011	North America / Pennsylvania	USA	41.9966	-80.03055	Coyote (Canis latrans)
TgCoPa04	PEN07	2011	North America / Pennsylvania	USA	40.9666	-78.435	Coyote (Canis latrans)
TgCoPa02	PEN08	2012	North America / Pennsylvania	USA	41.44083	-76.50861	Coyote (Canis latrans)
TgCoPa03	PEN09	2011	North America / Pennsylvania	USA	41.50055	-76.00833	Coyote (Canis latrans)
TgWTDPa06	PEN10	2011	North America / Pennsylvania	USA	40.043888	-75.19666	White-tailed deer (Odocoileus virginianus)
TgFoxPa06	PEN11	2011	North America / Pennsylvania	USA	39.8697222	-77.222	Red fox (Vulpes vulpes)
TgGoosePa01	PEN12	2011	North America / Pennsylvania	USA	40.043888	-75.19666	Canada Goose (Branta canadensis)
TgJagBr01	BRN01	2009	South America / Northeast Brazil	Brazil	-8,05	-34,88	Jaguarundi (Puma yagouaroundi)
TgCkBr181	BRN02	2007	South America / Northeast Brazil	Brazil	-4,97	-39,01	Chicken (Gallus gallus domesticus)
TgCkBr172	BRN03	2007	South America / Northeast Brazil	Brazil	-3,74	-43,36	Chicken (Gallus gallus domesticus)
TgDonkeyBr01	BRN04	2011	South America / Northeast Brazil	Brazil	-5,18	-37,34	Donkey (Equus africanus asinus)
TgOncBr01	BRN05	2010	South America / Northeast Brazil	Brazil	-8,05	-34,88	Oncilla (Leopardus tigrinus)
TgCkBr180	BRN06	2007	South America / Northeast Brazil	Brazil	-4,97	-39,01	Chicken (Gallus gallus domesticus)
TgCkBr167	BRN07	2007	South America / Northeast Brazil	Brazil	-6,70333	-36,94444	Chicken (Gallus gallus domesticus)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgCkBr176	BRN08	2007	South America / Northeast Brazil	Brazil	-11,09167	-40,43306	Chicken (Gallus gallus domesticus)
TgCkBr179	BRN09	2007	South America / Northeast Brazil	Brazil	-4,13	-38,23	Chicken (Gallus gallus domesticus)
TgCkBr183	BRN10	2007	South America / Northeast Brazil	Brazil	-10,68	-37,42	Chicken (Gallus gallus domesticus)
TgCkBr184	BRN11	2007	South America / Northeast Brazil	Brazil	-10,18	-36,84	Chicken (Gallus gallus domesticus)
TgCkBr170	BRN12	2007	South America / Northeast Brazil	Brazil	-5,07972	-37,61667	Chicken (Gallus gallus domesticus)
TgCpBr03	BRS01	2005	South America / Southeast Brazil	Brazil	-20,89944	-51,37833	Capybara (Hydrochaeris hydrochaeris)
TgCatBr52	BRS02	2003	South America / Southeast Brazil	Brazil	-22,19	-46,74	Cat (Felis silvestris catus)
TgCpBr30	BRS03	2005	South America / Southeast Brazil	Brazil	-23,54	-46,63	Capybara (Hydrochaeris hydrochaeris)
TgCpBr25	BRS04	2005	South America / Southeast Brazil	Brazil	-23,54	-46,63	Capybara (Hydrochaeris hydrochaeris)
TgCpBr18	BRS05	2005	South America / Southeast Brazil	Brazil	-20,47	-49,77	Capybara (Hydrochaeris hydrochaeris)
TgCatBr60	BRS06	2003	South America / Southeast Brazil	Brazil	-20,32139	-48,31222	Cat (Felis silvestris catus)
TgCpBr20	BRS07	2005	South America / Southeast Brazil	Brazil	-20,47	-49,77	Capybara (Hydrochaeris hydrochaeris)
TgShBr16	BRS08	2005	South America / Southeast Brazil	Brazil	-23,11	-47,71	Sheep (Ovis aries)
TgCpBr09	BRS09	2005	South America / Southeast Brazil	Brazil	-22,48	-47,45	Capybara (Hydrochaeris hydrochaeris)
TgCpBr33	BRS10	2005	South America / Southeast Brazil	Brazil	-21,22	-50,86	Capybara (Hydrochaeris hydrochaeris)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgCpBr27	BRS11	2005	South America / Southeast Brazil	Brazil	-23,54	-46,63	Capybara (Hydrochaeris hydrochaeris)
TgCatBr69	BRS12	2003	South America / Southeast Brazil	Brazil	-21,99	-47,42	Cat (Felis silvestris catus)
TgCatBr42	BRS13	2003	South America / Southeast Brazil	Brazil	-20,71361	-48,53556	Cat (Felis silvestris catus)
TgCatBr71	BRS14	2003	South America / Southeast Brazil	Brazil	-21,99	-47,42	Cat (Felis silvestris catus)
TgCatBr55	BRS15	2003	South America / Southeast Brazil	Brazil	-22,19	-46,74	Cat (Felis silvestris catus)
TgH18007	GUY01	2003	South America / Guiana shield	French Guiana	5.15	-53.766666	Human / Amazonian toxoplasmosis
GUY-GAL-VIT-01-(TgA18005)	GUY02	2009	South America / Guiana shield	French Guiana	4.891341	-52.348726	Grison (Galictis vittata)
GUY-JAG1-(TgA18001)	GUY03	2004	South America / Guiana shield	French Guiana	4.309389	-52.378005	Jaguar (Panthera onca)
TgH18006	GUY04	2003	South America / Guiana shield	French Guiana	4.3577	-53.23055	Human / Amazonian toxoplasmosis
TgH00002	GUY05	1991	South America / Guiana shield	French Guiana	4.3577	-53.23055	Human / Amazonian toxoplasmosis
TgH18008	GUY06	2004	South America / Guiana shield	Suriname	5.460021	-54.099640	Human / Amazonian toxoplasmosis
GUY-CAN-FAM-07-(TgA18006)	GUY07	2009	South America / Guiana shield	French Guiana	4.726325	-52.326405	Dog (Canis lupus familiaris)
TgH18009	GUY08	2004	South America / Guiana shield	French Guiana	5.271795	-54.258728	Human / Amazonian toxoplasmosis
TgH18013	GUY09	2004	South America / Guiana shield	French Guiana	4.82274	-53.276409	Human / Amazonian toxoplasmosis
TgH00009	GUY10	1997	South America / Guiana shield	French Guiana	4.3577	-53.23055	Human / Amazonian toxoplasmosis

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgH42001	GUY11	2010	South America / Guiana shield	French Guiana	4.766	-52.4866	Human / Amazonian toxoplasmosis
TgH18028	GUY12	2007	South America / Guiana shield	French Guiana	3.433378	-53.966817	Human / Amazonian toxoplasmosis
TgH18054	GUY13	2012	South America / Guiana shield	French Guiana	4.32055	-52.139166	Human / Amazonian toxoplasmosis
TgH18021	GUY14	2006	South America / Guiana shield	Suriname	4.857601	-54.473208	Human / Amazonian toxoplasmosis
TgH18057	GUY15	2014	South America / Guiana shield	French Guiana	3.8977	-51.8355	Human / Amazonian toxoplasmosis
TgH18049	GUY16	2011	South America / Guiana shield	French Guiana	3.6711	-54.0033	Human / Amazonian toxoplasmosis
TgH19006	GUY17	2012	South America / Guiana shield	French Guiana	3.6711	-54.0033	Human / Amazonian toxoplasmosis
TgCatCHn01	CHI01	2014	Asia / China	China	35.410277	113.9747222	Cat (Felis silvestris catus)
TgCtxz03	CHI02	*<2013*	Asia / China	China	34.20333	117.28444	Cat (Felis silvestris catus)
TgCtxz01	CHI03	*<2013*	Asia / China	China	34.20333	117.28444	Cat (Felis silvestris catus)
TgCatCHn03	CHI04	2014	Asia / China	China	34.74722	113.6477	Cat (Felis silvestris catus)
TgCatBj02	CHI05	2010	Asia / China	China	39.96	116.29805	Cat (Felis silvestris catus)
TgCatCZg03	CHI06	2014	Asia / China	China	28.979166	120.13166	Cat (Felis silvestris catus)
TgCtwh03	CHI07	<2011	Asia / China	China	30.5525	114.27139	Cat (Felis silvestris catus)
TgCtwh06	CHI08	<2011	Asia / China	China	30.5525	114.27139	Cat (Felis silvestris catus)
TgCtxz06	CHI09	*<2013*	Asia / China	China	34.20333	117.28444	Cat (Felis silvestris catus)
TgCatCZg05	CHI10	2014	Asia / China	China	28.979166	120.13166	Cat (Felis silvestris catus)
TgCtxz02	CHI11	*<2013*	Asia / China	China	34.20333	117.28444	Cat (Felis silvestris catus)
TgCatCZg02	CHI12	2014	Asia / China	China	28.979166	120.13166	Cat (Felis silvestris catus)
TgCtwh14	CHI13	*<2013*	Asia / China	China	30.5525	114.27139	Cat (Felis silvestris catus)
TgGZ02	CHI14	2012	Asia / China	China	26.65056	106.62444	Pig (Sus scrofa domesticus)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgCatTr_İzmir06	TUR01	2011	Asia / Turkey	Turkey	38.41388	27.14388	Cat (Felis silvestris catus)
TgCatTr_İzmir22	TUR02	2012	Asia / Turkey	Turkey	38.4777	27.097222	Cat (Felis silvestris catus)
TgCatTr_İzmir09	TUR03	2011	Asia / Turkey	Turkey	38.41388	27.14388	Cat (Felis silvestris catus)
TgCatTr_İzmir18	TUR04	2012	Asia / Turkey	Turkey	38.4777	27.097222	Cat (Felis silvestris catus)
TgCatTr_İzmir12	TUR05	2012	Asia / Turkey	Turkey	38.4777	27.097222	Cat (Felis silvestris catus)
TgCatTr_İzmir11	TUR06	2012	Asia / Turkey	Turkey	38.41388	27.14388	Cat (Felis silvestris catus)
TgCatTr_İzmir02	TUR07	2011	Asia / Turkey	Turkey	38.402499	27.03861	Cat (Felis silvestris catus)
TgCatTr_İzmir03	TUR08	2011	Asia / Turkey	Turkey	38.41388	27.14388	Cat (Felis silvestris catus)
TgCatTr_İzmir19	TUR09	2012	Asia / Turkey	Turkey	38.4777	27.097222	Cat (Felis silvestris catus)
TgCatTr_İzmir20	TUR10	2012	Asia / Turkey	Turkey	38.4777	27.097222	Cat (Felis silvestris catus)
Ankara-(TgH102001)	TUR11	1972	Asia / Turkey	Turkey	39.936944	32.846944	Human (congenital)
TgPiPr02	POR01	2005	Europe / Portugal	Portugal	41.83083	-7.00917	Pig (Sus scrofa domesticus)
TgCkPr16	POR02	2005	Europe / Portugal	Portugal	41.53333	-8.61667	Chicken (Gallus gallus domesticus)
TgCkPr02	POR03	2005	Europe / Portugal	Portugal	41.28056	-8.66833	Chicken (Gallus gallus domesticus)
TgCkPr01	POR04	2005	Europe / Portugal	Portugal	40.25417	-8.75472	Chicken (Gallus gallus domesticus)
TgPiPr14	POR05	2005	Europe / Portugal	Portugal	41.83083	-7.00917	Pig (Sus scrofa domesticus)
TgCkPr04	POR06	2005	Europe / Portugal	Portugal	41.28056	-8.66833	Chicken (Gallus gallus domesticus)
TgCkPr03	POR07	2005	Europe / Portugal	Portugal	41.28056	-8.66833	Chicken (Gallus gallus domesticus)
TgPiPr07	POR08	2005	Europe / Portugal	Portugal	41.83083	-7.00917	Pig (Sus scrofa domesticus)
TgPiPr05	POR09	2005	Europe / Portugal	Portugal	41.83083	-7.00917	Pig (Sus scrofa domesticus)
TgPiPr13	POR10	2005	Europe / Portugal	Portugal	41.83083	-7.00917	Pig (Sus scrofa domesticus)
TgCkPr11	POR11	2005	Europe / Portugal	Portugal	40.68917	-8.47972	Chicken (Gallus gallus domesticus)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgPiPr09	POR12	2005	Europe / Portugal	Portugal	41.83083	-7.00917	Pig (Sus scrofa domesticus)
FR-OVI-ARI056-(TgA32122)	FRA01	2007	Europe / France	France	43.268055	-0.37944	Sheep (Ovis aries)
FR-OVI-ARI049-(TgA32115)	FRA02	2007	Europe / France	France	43.268055	-0.37944	Sheep (Ovis aries)
FR-OVI-ARI026-(TgA32092)	FRA03	2007	Europe / France	France	43.23388	0.14611	Sheep (Ovis aries)
FR-OVI-ARI021-(TgA32088)	FRA04	2007	Europe / France	France	44.158055	6.084166	Sheep (Ovis aries)
FR-OVI-ARI050-(TgA32116)	FRA05	2007	Europe / France	France	43.268055	-0.37944	Sheep (Ovis aries)
FR-OVI-ARI042-(TgA32108)	FRA06	2007	Europe / France	France	44.4422	1.546944	Sheep (Ovis aries)
FR-OVI-ARI029-(TgA32095)	FRA07	2007	Europe / France	France	44.405	4.9388	Sheep (Ovis aries)
FR-OVI-ARI027-(TgA32093)	FRA08	2007	Europe / France	France	44.405	4.9388	Sheep (Ovis aries)
FR-OVI-ARI033-(TgA32099)	FRA09	2007	Europe / France	France	46.007222	0.208333	Sheep (Ovis aries)
FR-OVI-ARI025-(TgA32091)	FRA10	2007	Europe / France	France	43.23388	0.14611	Sheep (Ovis aries)
FR-OVI-ARI043-(TgA32109)	FRA11	2007	Europe / France	France	44.4422	1.546944	Sheep (Ovis aries)
FR-OVI-ARI022-(TgA32129)	FRA12	2007	Europe / France	France	44.158055	6.084166	Sheep (Ovis aries)
TgCkAt14	AUS01	2004	Europe / Austria	Austria	47.33417	15.95806	Chicken (Gallus gallus domesticus)
TgCkAt30	AUS02	2004	Europe / Austria	Austria	46.72583	15.23806	Chicken (Gallus gallus domesticus)
TgCkAt06	AUS03	2004	Europe / Austria	Austria	46.80667	15.31444	Chicken (Gallus gallus domesticus)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Name in this study	Year	Population	Country	Latitude (decimal degree)	Longitude (decimal degree)	Host
TgCkAt32	AUS04	2004	Europe / Austria	Austria	46.72583	15.23806	Chicken (Gallus gallus domesticus)
TgCkAt46	AUS05	2004	Europe / Austria	Austria	46.72583	15.23806	Chicken (Gallus gallus domesticus)
TgCkAt07	AUS06	2004	Europe / Austria	Austria	46.80667	15.31444	Chicken (Gallus gallus domesticus)
TgCkAt17	AUS07	2004	Europe / Austria	Austria	46.82417	15.3175	Chicken (Gallus gallus domesticus)
TgCkAt27	AUS08	2004	Europe / Austria	Austria	46.72583	15.23806	Chicken (Gallus gallus domesticus)
TgCkAt20	AUS09	2004	Europe / Austria	Austria	46.82417	15.3175	Chicken (Gallus gallus domesticus)
TgCkAt08	AUS10	2004	Europe / Austria	Austria	46.87194	15.90278	Chicken (Gallus gallus domesticus)

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers														Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
TgSpEt23	289	242	205	160	336	165	278	356	213	190	147	111	271	089	312	[1, 2]
TgGtEt05	289	242	205	160	336	165	278	356	213	190	147	113	267	089	312	[1, 2]
TgSpEt18	289	242	207	158	336	169	274	356	211	176	140	109	267	091	308	[1, 2]
TgSpEt25	289	242	207	158	336	169	274	356	211	176	140	121	267	091	308	[1, 2]
TgGtEt08	289	242	207	158	336	169	274	356	211	178	142	109	285	085	310	[1, 2]
TgSpEt06	289	242	207	158	336	169	274	356	213	176	140	115	265	117	308	[1, 2]
TgSpEt08	289	242	207	158	336	169	274	356	213	178	140	115	265	089	308	[1, 2]
TgGtEt19	289	242	207	158	336	169	274	356	215	176	138	117	267	099	312	[1, 2]
TgGtEt01	289	242	207	158	336	169	274	356	225	172	142	117	259	093	308	[1, 2]
TgGtEt14	289	242	207	158	336	169	274	356	233	176	140	113	269	097	312	[1, 2]
TgGtEt02	289	242	207	158	336	169	274	356	235	176	138	115	279	099	312	[1, 2]
TgGtEt04	289	242	207	158	336	169	274	356	237	176	140	113	273	093	312	[1, 2]
TgSpEt26	289	242	207	158	336	169	274	356	239	176	138	115	281	097	310	[1, 2]
GAB1-GAL-DOM06-(TgA105015)	289	242	205	160	336	165	278	356	211	190	147	111	267	085	312	[3]
GAB1-GAL-DOM11-(TgA105016)	289	242	205	160	336	165	278	356	213	190	145	111	273	089	312	[3]
GAB1-CAP-AEG06-(TgA105011)	289	242	205	160	336	165	278	356	213	190	149	111	267	089	312	[3]
GAB1-GAL-DOM13-(TgA105018)	289	242	205	160	336	165	278	356	213	190	149	111	267	089	312	[3]
GAB2-GAL-DOM06-(TgA105040)	289	242	207	160	336	165	278	356	223	190	147	111	261	103	316	[3]
GAB4-GAL-DOM01-(TgA105053)	291	242	205	160	336	165	278	356	213	190	145	111	269	089	306	[3]

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers															Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
GAB2-GAL-DOM02-(TgA105004)	291	242	207	160	342	165	278	354	223	166	142	111	277	097	310	[3]
GAB2-GAL-DOM01-(TgA105037)	291	242	207	160	342	165	278	354	223	166	142	111	279	097	310	[3]
GAB1-GAL-DOM10-(TgA105003)	291	242	207	160	342	165	278	354	225	166	142	111	275	097	310	[3]
GAB3-GAL-DOM08-(TgA105052)	291	242	207	160	342	165	278	354	227	166	142	111	275	097	310	[3]
GAB5-GAL-DOM03-(TgA105055)	291	242	207	160	342	165	278	356	225	166	142	111	271	093	310	[3]
GAB3-GAL-DOM02-(TgA105001)	291	248	205	160	342	165	274	354	223	166	147	111	269	089	306	[3]
GAB3-GAL-DOM11-(TgA105043)	291	248	205	160	342	165	274	354	223	166	147	111	273	087	304	[3]
GAB5-GAL-DOM01-(TgA105002)	291	248	205	160	342	165	274	354	231	166	149	111	277	087	306	[3]
TONT	289	242	205	162	336	165	278	356	213	164	142	111	279	087	312	[4]
TgCatPr09	289	242	205	164	342	165	278	356	213	172	147	111	267	091	306	[5]
TgH16001	289	242	207	158	336	169	274	356	213	174	147	119	271	089	335	[6]
TgH40001	289	242	207	162	336	169	274	356	237	166	142	111	281	093	312	This study
TgCatPr07	289	246	207	162	336	171	274	356	227	166	147	113	267	097	312	[5]
TgCatPr08	289	250	209	160	336	165	278	358	213	166	142	111	265	087	310	[5]

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers														Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
ENVL-2002-MAC	291	242	205	160	336	165	278	356	213	190	142	111	277	087	312	[7]
TgH40002	291	242	205	162	336	165	278	356	211	164	147	109	277	087	312	This study
TgH16010	291	242	205	162	336	165	278	356	213	164	142	111	283	085	312	This study
TgCatStk07a	291	242	205	162	342	165	278	356	211	164	142	109	277	087	312	[8]
TgH13006	291	242	205	162	342	165	278	356	213	164	142	109	277	087	312	[6]
TgCatPr06	291	246	209	162	336	169	274	356	213	168	142	113	267	087	306	[5]
TgCatPr10	291	248	209	160	336	169	274	356	213	166	145	111	263	087	306	[5]
TgH16012	289	242	205	162	336	165	278	356	211	164	142	109	277	085	312	This study
TgWolfMN20	289	242	205	160	336	165	278	356	213	190	153	111	267	089	312	[9]
TgWtdUs10	289	242	205	160	336	165	278	356	215	190	151	111	259	093	312	[10]
TgWtdUs08	289	242	207	160	336	169	274	356	227	190	142	111	269	103	310	[10]
TgWolfMN11	289	242	209	160	336	169	274	362	215	170	147	113	281	091	318	[9]
TgWolfMN29	289	242	209	160	336	169	274	362	215	170	151	113	291	101	320	[9]
TgWolfMN19	289	242	211	160	336	169	274	360	227	170	NA	127	283	111	324	[9]
TgWolfMN13	289	242	211	160	336	169	274	362	215	170	151	113	279	101	314	[9]
TgWolfMN28	289	242	211	160	336	169	274	362	225	178	NA	113	273	113	314	[9]
TgWolfMN26	289	242	211	160	336	169	274	362	239	172	145	111	283	111	316	[9]
TgWolfMN12	289	242	211	162	336	169	274	362	217	170	149	111	295	093	316	[9]
TgWolfMN27	289	242	211	162	336	169	274	362	223	170	153	113	295	105	314	[9]
TgWolfMN25	289	242	211	162	336	169	274	362	225	170	151	111	279	109	314	[9]
TgFoxPa03	289	242	205	160	336	165	278	356	213	190	142	111	269	089	312	[11]
TgCoPa08	289	242	207	158	336	169	274	356	213	174	142	123	271	097	310	[11]
TgCoPa05	289	242	207	158	336	169	274	356	215	174	140	119	261	099	308	[11]
TgCoPa01	289	242	207	158	336	169	274	356	235	174	142	121	261	093	310	[11]
TgFoxPa10	289	242	207	162	336	169	274	356	213	170	142	111	301	103	316	[11]

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers														Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
TgCoPa07	289	242	211	160	336	169	274	356	213	178	145	111	267	109	316	[11]
TgCoPa04	289	242	211	160	336	169	274	362	213	172	149	113	285	095	331	[11]
TgCoPa02	289	242	211	160	336	169	274	362	225	170	166	113	295	093	310	[11]
TgCoPa03	289	242	211	162	336	169	274	360	237	172	151	111	269	117	316	[11]
TgWTDPa06	289	242	213	160	336	169	274	356	215	182	145	111	269	101	316	[11]
TgFoxPa06	289	248	209	160	336	165	278	356	213	166	149	107	265	087	306	[11]
TgGoosePa01	291	242	209	160	336	165	278	356	209	190	149	111	263	087	306	[11]
TgJagBr01	289	242	205	162	344	165	278	358	235	164	138	111	263	091	312	[12]
TgCkBr181	289	242	205	164	344	165	278	356	223	172	140	111	267	087	312	[13, 14]
TgCkBr172	289	242	207	160	348	169	278	356	213	196	149	111	265	087	308	[13, 14]
TgDonkeyBr01	291	242	205	156	348	169	278	358	213	170	140	111	263	121	314	[15]
TgOncBr01	291	242	205	160	342	165	278	356	227	166	147	111	281	093	308	[16]
TgCkBr180	291	242	205	162	342	165	278	356	213	164	142	109	277	087	312	[13, 14]
TgCkBr167	291	242	205	162	342	165	278	356	213	164	142	109	281	087	314	[13, 14]
TgCkBr176	291	242	205	162	342	165	278	356	213	164	142	109	277	087	312	[13, 14]
TgCkBr179	291	242	205	162	342	165	278	356	213	164	142	109	277	087	312	[13, 14]
TgCkBr183	291	242	205	162	342	165	278	356	213	164	142	109	277	087	312	[13, 14]
TgCkBr184	291	242	205	162	342	165	278	356	213	164	142	109	277	087	312	[13, 14]
TgCkBr170	291	242	205	162	342	165	278	356	213	164	142	109	277	089	312	[13, 14]
TgCpBr03	289	238	205	162	350	169	276	356	215	172	140	117	322	109	320	[17, 18]
TgCatBr52	289	242	205	160	342	165	278	358	231	164	145	111	316	089	308	[19]
TgCpBr30	289	242	205	160	344	165	278	356	227	190	142	111	293	091	333	[17, 18]
TgCpBr25	289	242	205	160	344	165	278	356	227	190	142	111	295	091	331	[17, 18]
TgCpBr18	289	242	205	160	348	165	278	356	213	190	142	111	261	127	310	[17, 18]
TgCatBr60	289	242	205	160	348	165	278	356	213	190	142	111	263	107	312	[19]
TgCpBr20	289	242	205	160	348	165	278	356	213	190	142	111	263	123	310	[17, 18]
TgShBr16	289	242	207	160	338	165	278	356	227	190	136	105	263	095	310	[20, 21]

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers															Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
TgCpBr09	291	242	205	160	342	165	278	354	229	166	140	111	271	091	308	[17, 18]
TgCpBr33	291	242	205	160	342	165	278	356	229	174	140	111	265	091	314	[17, 18]
TgCpBr27	291	242	205	160	360	165	278	356	233	166	149	111	263	091	308	[17, 18]
TgCatBr69	291	242	205	162	342	165	278	356	233	166	166	111	265	091	341	[19]
TgCatBr42	291	248	205	160	342	165	274	354	229	166	147	111	273	089	308	[19]
TgCatBr71	291	248	205	160	342	165	274	354	233	166	149	111	271	089	308	[19]
TgCatBr55	291	248	207	160	342	169	272	356	249	166	145	111	267	089	312	[19]
TgH18007	289	242	203	158	344	171	272	356	209	182	149	121	265	089	317	[7]
GUY-GAL-VIT-01-(TgA18005)	289	242	203	160	336	165	272	356	213	176	140	105	279	087	312	[22]
GUY-JAG1-(TgA18001)	289	242	203	160	336	169	274	358	209	176	142	111	257	107	341	[23]
TgH18006	289	242	203	160	344	165	276	356	211	168	142	109	271	087	320	[24]
TgH00002	289	242	205	170	360	167	274	356	223	190	142	109	259	085	312	[25]
TgH18008	289	246	203	158	338	167	276	354	213	168	138	111	281	093	318	[26]
GUY-CAN-FAM-07-(TgA18006)	289	246	209	160	342	173	272	356	239	168	138	113	261	087	312	[22]
TgH18009	291	242	203	160	338	167	274	356	209	188	138	115	263	091	312	This study
TgH18013	291	242	203	160	346	167	272	356	217	170	147	127	257	085	310	This study
TgH00009	291	242	203	162	344	167	276	356	217	170	142	113	277	091	308	[27, 7]
TgH42001	291	246	203	158	346	169	272	356	229	172	142	105	261	105	316	[28]
TgH18028	291	246	203	160	326	167	276	356	211	178	147	111	259	087	306	This study
TgH18054	291	246	203	162	348	165	272	356	221	170	142	107	261	091	318	This study
TgH18021	291	246	205	166	334	167	272	356	213	174	151	107	267	087	325	This study

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers														Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
TgH18057	291	248	209	160	342	165	278	354	209	166	140	113	277	095	304	This study
TgH18049	293	242	203	162	344	167	274	358	217	176	138	115	273	085	316	This study
TgH19006	293	246	203	162	344	171	272	356	211	182	138	105	267	091	302	This study
TgCatCHn01	289	242	205	160	336	165	278	356	213	190	147	111	265	089	312	[29]
TgCtxz03	289	242	207	160	346	169	274	356	213	174	140	129	259	107	306	[30]
TgCtxz01	291	248	209	160	342	169	274	358	209	166	145	119	265	089	306	[30]
TgCatCHn03	293	242	211	160	336	169	274	354	215	172	130	111	277	087	308	[29]
TgCatBj02	293	242	211	160	336	169	274	354	215	172	130	111	283	093	306	[31]
TgCatCZg03	293	242	211	160	336	169	274	354	215	172	130	111	289	097	306	[29]
TgCtwh03	293	242	211	160	336	169	274	354	215	172	145	111	273	093	306	[32]
TgCtwh06	293	242	211	160	336	169	274	354	215	172	145	111	287	093	306	[32]
TgCtxz06	293	242	211	160	336	169	274	354	215	172	145	111	297	095	306	[30]
TgCatCZg05	293	242	211	160	336	169	274	354	215	172	145	111	303	093	306	[29]
TgCtxz02	293	242	211	160	336	169	274	354	215	172	145	113	285	091	308	[30]
TgCatCZg02	293	242	211	160	336	169	274	354	215	172	145	125	287	091	308	[29]
TgCtwh14	293	242	211	160	336	169	274	354	215	172	145	125	297	091	308	[30]
TgGZ02	293	242	211	160	336	169	274	354	219	172	130	111	279	093	308	[33]
TgCatTr_İzmir06	289	242	207	158	336	169	274	356	211	176	140	123	275	085	314	[34]
TgCatTr_İzmir22	289	242	207	158	336	169	274	356	213	172	145	111	267	097	310	[34]
TgCatTr_İzmir09	289	242	207	158	336	169	274	356	213	178	140	119	295	101	310	[34]
TgCatTr_İzmir18	289	242	207	158	336	169	274	356	215	174	140	125	261	087	314	[34]
TgCatTr_İzmir12	289	242	207	158	336	169	274	356	215	178	138	109	297	109	310	[34]
TgCatTr_İzmir11	289	242	207	158	336	169	274	356	223	176	142	127	277	085	314	[34]
TgCatTr_İzmir02	289	242	207	158	336	169	274	356	227	172	145	131	263	099	310	[34]
TgCatTr_İzmir03	289	242	207	158	336	169	274	356	227	176	140	115	279	095	312	[34]

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers															Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
TgCatTr_İzmir19	289	242	207	158	336	169	274	356	231	174	142	111	263	097	314	[34]
TgCatTr_İzmir20	289	242	207	158	336	169	274	356	241	176	140	113	271	089	310	[34]
Ankara-(TgH102001)	291	248	205	160	342	165	274	354	227	166	147	111	297	091	310	[35]
TgPiPr02	289	242	205	160	336	165	278	356	209	192	145	111	269	089	312	[36]
TgCkPr16	289	242	205	160	336	165	278	356	213	188	140	119	265	089	312	[37, 38]
TgCkPr02	289	242	205	160	336	165	278	356	213	190	147	111	263	089	312	[37, 38]
TgCkPr01	289	242	205	160	336	165	278	356	213	190	147	111	267	087	312	[37, 38]
TgPiPr14	289	242	205	160	336	165	278	356	213	190	149	111	267	089	312	[36]
TgCkPr04	289	242	205	160	336	165	278	356	215	190	147	111	267	085	312	[37, 38]
TgCkPr03	289	242	207	158	336	169	274	356	213	176	140	115	265	109	316	[37, 38]
TgPiPr07	289	242	207	158	336	169	274	356	213	176	140	115	273	093	316	[36]
TgPiPr05	289	242	207	158	336	169	274	356	213	176	145	113	259	093	310	[36]
TgPiPr13	289	242	207	158	336	169	274	356	219	174	140	127	269	091	308	[36]
TgCkPr11	289	242	207	158	336	169	274	356	223	174	140	115	273	093	316	[37, 38]
TgPiPr09	289	242	207	158	336	169	274	356	225	174	140	119	265	101	308	[36]
FR-OVI-ARI056-(TgA32122)	289	242	205	160	336	165	278	356	213	190	149	111	267	*089*	ND	[39]
FR-OVI-ARI049-(TgA32115)	289	242	207	158	336	169	274	356	211	176	140	113	261	097	312	[39]
FR-OVI-ARI026-(TgA32092)	289	242	207	158	336	169	274	356	211	176	140	113	275	105	312	[39]
FR-OVI-ARI021-(TgA32088)	289	242	207	158	336	169	274	356	213	174	140	109	259	089	308	[39]
FR-OVI-ARI050-(TgA32116)	289	242	207	158	336	169	274	356	213	174	140	135	267	109	312	[39]
FR-OVI-ARI042-(TgA32108)	289	242	207	158	336	169	274	356	215	176	140	119	273	091	312	[39]

BY-NC-ND

Isolate (Toxoplasma BRC code number)	Microsatellite markers														Ref.
	TUB2	*W35*	*TgM-A*	*B18*	*B17*	M33	MIV.1	MXI.1	*M48*	*M102*	*N60*	*N82*	*AA*	*N61*	*N83*	
FR-OVI-ARI029-(TgA32095)	289	242	207	158	336	169	274	356	217	172	140	115	263	099	310	[39]
FR-OVI-ARI027-(TgA32093)	289	242	207	158	336	169	274	356	217	178	140	111	279	095	312	[39]
FR-OVI-ARI033-(TgA32099)	289	242	207	158	336	169	274	356	227	174	145	111	263	101	314	[39]
FR-OVI-ARI025-(TgA32091)	289	242	207	158	336	169	274	356	227	174	145	123	299	091	312	[39]
FR-OVI-ARI043-(TgA32109)	289	242	207	158	336	169	274	356	227	176	140	115	267	093	310	[39]
FR-OVI-ARI022-(TgA32129)	289	242	207	158	336	169	274	356	227	176	140	119	273	099	314	[39]
TgCkAt14	289	242	207	158	336	169	274	356	211	174	142	109	263	103	310	[40, 38]
TgCkAt30	289	242	207	158	336	169	274	356	213	174	138	133	259	105	310	[40, 38]
TgCkAt06	289	242	207	158	336	169	274	356	219	174	142	113	259	095	312	[40, 38]
TgCkAt32	289	242	207	158	336	169	274	356	213	174	138	129	259	111	310	[40, 38]
TgCkAt46	289	242	207	158	336	169	274	356	225	176	140	115	277	099	312	[40, 38]
TgCkAt07	289	242	207	158	336	169	274	356	223	176	145	119	265	087	310	[40, 38]
TgCkAt17	289	242	207	158	336	169	274	356	229	176	138	115	261	087	310	[40, 38]
TgCkAt27	289	242	207	158	336	169	274	356	221	174	142	125	259	091	312	[40, 38]
TgCkAt20	289	242	207	158	336	169	274	356	213	174	138	129	259	109	310	[40, 38]
TgCkAt08	289	242	207	158	336	169	274	356	211	174	142	109	261	101	310	[40, 38]

BY-NC-ND

Ref. References:

1. Gebremedhin E, Endrias G, Abebe A, Tesfaye T, Getachew T, Girmay M, et al. Seroepidemiological study of ovine toxoplasmosis in East and West Shewa Zones of Oromia Regional State, Central Ethiopia. BMC Vet Res. 2013;9: 117.

2. Gebremedhin EZ, Abdurahaman M, Tessema TS, Tilahun G, Cox E, Goddeeris B, et al. Isolation and genotyping of viable *Toxoplasma gondii* from sheep and goats in Ethiopia destined for human consumption. Parasit Vectors. 2014;7: 425.

3. Mercier A, Devillard S, Ngoubangoye B, Bonnabau H, Bañuls A-L, Durand P, et al. Additional haplogroups of *Toxoplasma gondii* out of Africa: population structure and mouse-virulence of strains from Gabon. PLoS Negl Trop Dis. 2010;4: e876.

4. Ajzenberg D, Daniel A, Nadine C, Luc P, Marie-Hélène B, Philippe T, et al. Genotype of 86 *Toxoplasma gondii* Isolates Associated with Human Congenital Toxoplasmosis, and Correlation with Clinical Findings. J Infect Dis. 2002;186: 684–689.

5. Dubey JP, López-Torres HY, Sundar N, Velmurugan GV, Ajzenberg D, Kwok OCH, et al.Mouse-virulent *Toxoplasma gondii* isolated from feral cats on Mona Island, Puerto Rico. J Parasitol. 2007;93: 1365–1369.

6. Ajzenberg D, Daniel A, Hélène Y, Pierre M, Luc P, Frédéric D, et al. Genotype of 88 *Toxoplasma gondii* Isolates Associated with Toxoplasmosis in Immunocompromised Patients and Correlation with Clinical Findings. J Infect Dis. 2009;199: 1155–1167.

7. Ajzenberg D, Bañuls AL, Su C, Dumètre A, Demar M, Carme B, et al. Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. Int J Parasitol. 2004;34: 1185–1196.

8. Dubey JP, L. M, D. M, N. S, Velmurugan GV, Kwok OCH, et al. Isolation and characterization of viable *Toxoplasma gondii* isolates revealed possible high frequency of mixed infection in feral cats (Felis domesticus) from St Kitts, West Indies. Parasitology. 2009;136: 589.

9. Dubey JP, Choudhary S, Ferreira LR, Kwok OCH, Butler E, Carstensen M, et al. Isolation and RFLP genotyping of *Toxoplasma gondii* from the gray wolf (Canis lupus). Vet Parasitol. 2013;197: 685–690.

10. Dubey JP, Velmurugan GV, Ulrich V, Gill J, Carstensen M, Sundar N, et al. Transplacental toxoplasmosis in naturally-infected white-tailed deer: Isolation and genetic characterisation of *Toxoplasma gondii* from foetuses of different gestational ages. Int J Parasitol. 2008;38: 1057–1063.

11. Dubey JP, Van Why K, Verma SK, Choudhary S, Kwok OCH, Khan A, et al. Genotyping *Toxoplasma gondii* from wildlife in Pennsylvania and identification of natural recombinants virulent to mice. Vet Parasitol. 2014;200: 74–84.

12. Pena HFJ, Marvulo MFV, Horta MC, Silva MA, Silva JCR, Siqueira DB, et al. Isolation and genetic characterisation of *Toxoplasma gondii* from a red-handed howler monkey (Alouatta belzebul), a jaguarundi (Puma yagouaroundi), and a black-eared opossum (Didelphis aurita) from Brazil. Vet Parasitol. 2011;175: 377–381.

13. de Oliveira LN, Costa Junior LM, de Melo CF, Ramos Silva JC, Bevilaqua CML, Azevedo SS, et al. *Toxoplasma gondii* Isolates From Free-Range Chickens From the Northeast Region of Brazil. J Parasitol. 2009;95: 235–237.

BY-NC-ND

14. Dubey JP, Velmurugan GV, Chockalingam A, Pena HFJ, de Oliveira LN, Leifer CA, et al. Genetic diversity of *Toxoplasma gondii* isolates from chickens from Brazil. Vet Parasitol. 2008;157: 299–305.

15. Gennari SM, de O. Esmerini P, Lopes MG, Soares HS, Vitaliano SN, Cabral AD, et al. Occurrence of antibodies against *Toxoplasma gondii* and its isolation and genotyping in donkeys, mules, and horses in Brazil. Vet Parasitol. 2015;209: 129–132.

16. Vitaliano SN, Soares HS, Minervino AHH, Santos ALQ, Werther K, Marvulo MFV, et al. Genetic characterization of *Toxoplasma gondii* from Brazilian wildlife revealed abundant new genotypes. Int J Parasitol Parasites Wildl. 2014;3: 276–283.

17. Yai LEO, Ragozo AMA, Aguiar DM, Damaceno JT, Oliveira LN, Dubey JP, et al. Isolation of *Toxoplasma gondii* from Capybaras (Hydrochaeris hydrochaeris) from São Paulo State, Brazil. J Parasitol. 2008;94: 1060–1063.

18. Yai LEO, Ragozo AMA, Soares RM, Pena HFJ, Su C, Gennari SM. Genetic diversity among capybara (Hydrochaeris hydrochaeris) isolates of *Toxoplasma gondii* from Brazil. Vet Parasitol. 2009;162: 332–337.

19. Pena HFJ, Soares RM, Amaku M, Dubey JP, Gennari SM. *Toxoplasma gondii* infection in cats from São Paulo state, Brazil: Seroprevalence, oocyst shedding, isolation in mice, and biologic and molecular characterization. Res Vet Sci. 2006;81: 58–67.

20. Ragozo AMA, Yai LEO, Oliveira LN, Dias RA, Dubey JP, Gennari SM. Seroprevalence and Isolation of *Toxoplasma gondii* From Sheep From São Paulo State, Brazil. J Parasitol. 2008;94: 1259–1263.

21. Ragozo AMA, Pena HFJ, Yai LEO, Su C, Gennari SM. Genetic diversity among *Toxoplasma gondii* isolates of small ruminants from Brazil: Novel genotypes revealed. Vet Parasitol. 2010;170: 307–312.

22. Mercier A, Ajzenberg D, Devillard S, Demar MP, de Thoisy B, Bonnabau H, et al. Human impact on genetic diversity of *Toxoplasma gondii*: Example of the anthropized environment from French Guiana. Infect Genet Evol. 2011;11: 1378–1387.

23. Demar M, Ajzenberg D, Serrurier B, Dardé ML, Carme B. Atypical *Toxoplasma gondii* strain from a free-living jaguar (Panthera onca) in French Guiana. Am J Trop Med Hyg. 2008;78: 195–197.

24. Su C, Khan A, Zhou P, Majumdar D, Ajzenberg D, M.-L. D, et al. Globally diverse *Toxoplasma gondii* isolates comprise six major clades originating from a small number of distinct ancestral lineages. Proceedings of the National Academy of Sciences. 2012;109: 5844–5849.

25. Dardé ML, Villena I, Pinon JM, Beguinot I. Severe toxoplasmosis caused by a *Toxoplasma gondii* strain with a new isoenzyme type acquired in French Guyana. J Clin Microbiol. 1998;36: 324.

26. Demar M, Ajzenberg D, Maubon D, Djossou F, Panchoe D, Punwasi W, et al. Fatal Outbreak of Human Toxoplasmosis along the Maroni River: Epidemiological, Clinical, and Parasitological Aspects. Clin Infect Dis. 2007;45: e88–e95.

BY-NC-ND

27. Bossi P, Caumes E, Paris L, Dardé ML, Bricaire F. *Toxoplasma gondii*-associated Guillain-Barré syndrome in an immunocompetent patient. J Clin Microbiol. déc 1998;36(12):3724‑5.

28. Groh M, Matthieu G, Alexandra F, Isabelle V, Daniel A, Bernard C, et al. Acute Lung, Heart, Liver, and Pancreatic Involvements with Hyponatremia and Retinochoroiditis in a 33-Year-Old French Guianan Patient. PLoS Negl Trop Dis. 2012;6: e1802.

29. Yang Y, Yurong Y, Yuqing Y, Verma SK, Cassinelli ABM, Kwok OCH, et al. Isolation and genetic characterization of viable *Toxoplasma gondii* from tissues and feces of cats from the central region of China. Vet Parasitol. 2015;211: 283–288.

30. Wang L, Chen H, Liu D, Huo X, Gao J, Song X, et al. Genotypes and mouse virulence of *Toxoplasma gondii* isolates from animals and humans in China. PLoS One. 2013;8: e53483.

31. Qian W, Weifeng Q, Hui W, Chunlei S, Dan S, Xia C, et al. Isolation and characterization of *Toxoplasma gondii* strains from stray cats revealed a single genotype in Beijing, China. Vet Parasitol. 2012;187: 408–413.

32. Chen ZW, Gao JM, Huo XX, Wang L, Yu L, Halm-Lai F, et al. Genotyping of *Toxoplasma gondii* isolates from cats in different geographic regions of China. Vet Parasitol. 2011;183: 166–170.

33. Li Y-N, Nie X, Peng Q-Y, Mu X-Q, Zhang M, Tian M-Y, et al. Seroprevalence and genotype of *Toxoplasma gondii* in pigs, dogs and cats from Guizhou province, Southwest China. Parasit Vectors. 2015;8: 214.

34. Can H, Döşkaya M, Ajzenberg D, Özdemir HG, Caner A, İz SG, et al. Genetic characterization of *Toxoplasma gondii* isolates and toxoplasmosis seroprevalence in stray cats of İzmir, Turkey. PLoS One. 2014;9: e104930.

35. Döşkaya M, Mert D, Ayşe C, Daniel A, Aysu D, Marie-Laure D, et al. Isolation of *Toxoplasma gondii* strains similar to Africa 1 genotype in Turkey. Parasitol Int. 2013;62: 471–474.

36. Sousa S de, de Sousa S, Ajzenberg D, Nuno C, Lina F, da Costa JMC, et al. Biologic and molecular characterization of *Toxoplasma gondii* isolates from pigs from Portugal. Vet Parasitol. 2006;135: 133–136.

37. Dubey JP, Vianna MCB, Sousa S, Canada N, Meireles S, da Costa JMC, et al. Characterization of *Toxoplasma gondii* Isolates in Free-Range Chickens From Portugal. J Parasitol. 2006;92: 184–186.

38. Verma SK, D. A, A. R-S, C. SU, Dubey JP. Genetic characterization of *Toxoplasma gondii* isolates from Portugal, Austria and Israel reveals higher genetic variability within the type II lineage. Parasitology. 2015;142: 948–957.

39. Halos L, Thébault A, Aubert D, Thomas M, Perret C, Geers R, et al. An innovative survey underlining the significant level of contamination by *Toxoplasma gondii* of ovine meat consumed in France. Int J Parasitol. 2010;40: 193–200.

40. Dubey JP, Edelhofer R, Marcet P, Vianna MCB, Kwok OCH, Lehmann T. Genetic and biologic characteristics of *Toxoplasma gondii* infections in free-range chickens from Austria. Vet Parasitol. 2005;133: 299–306.

BY-NC-ND

Annexe 2.2. S2 Table : Polymorphism of different *Toxoplasma gondii* DNA regions screened in GenBank

Locus	Alignment size	Number of sequences	S/L	P	GenBank accession numbers
GRA6	212	145	18/212	85	AB235426-AB235435, AB471758, AF239284-AF239292, AJ635333-AJ635336, AY964058-AY964060, DQ187387, DQ512729, EF512218-EF512239, EF512241-EF512265, EF585704-EF585715, EU180619, EU180621, EU180623, EU180624, EU573185, JX044178-JX044180, JX044183-JX044189, JX044191, JX044192, JX044194-JX044196, JX044199, JX044201, JX044203-JX044211, JX044214-JX044225, JX044227-JX044234, JX044236, JX044237, JX044240-JX044243
SAG3	157	24	13/157	83	AF340228, AF340229, AY187280, EF585683-EF585694, JX218224, KC928250-KC928254, GT1 (ToxoDB), ME49(ToxoDB), VEG (ToxoDB)
ROP18	1673	45	137/1673	82	EF092842, GQ243205, GQ243206, GQ243208-GQ243216, JX045318-JX045320, JX045322-JX045329, JX045331-JX045338, JX045340-JX045345, JX045347-JX045354
UPRT intron 1	399	140	31/399	78	AY143142, AY143145, EU573186, JQ679758-JQ679781, JQ679783-JQ679895
GRA7	415	130	30/415	72	DQ459444-DQ459458, DQ465955, DQ473315, EF626965-EF626970, EF639857-EF639859, EU157141-EU157183, HQ852155-HQ852159, JX045574-JX045586, JX045588-JX045596, JX045600, JX045604-JX045615, JX045617-JX045632, JX045634-JX045638
UPRT intron 7	459	140	24/459	52	AY143148, AY143149, AY143151-AY143153, JQ679896-JQ679931, JQ679933-JQ680031
ROP16	2129	17	105/2129	49	AAYL02000151, GQ249081-GQ249096
GRA3	663	24	29/663	44	JX044153-JX044163, JX044165-JX044177
locus 5 Chr Ia	503	74	20/503	40	JX044476-JX044489, JX044491, JX044493-JX044543, JX044545-JX044552

BY-NC-ND

SAG2	351	42	14/351	40	AF249696-AF249698, AF357577-AF357582, AY187279, EU258524-EU258530, EU258532, EU258533, EU650329, EU650330, JX045470, JX045473-JX045477, JX045479-JX045491, JX045493
Hp intron 2	470	137	16/470	35	JQ679621-JQ679757
MIC2 intron 3	372	75	11/372	30	JX044322-JX044324, JX044326-JX044383, JX044386-JX044399
Actin intron	432	25	12/432	28	AF247381, AF247382, JX045422-JX045443, JX045445
locus 21 Chr. Ia	721	73	19/721	27	JX045166-JX045187, JX045190-JX045223, JX045225-JX045241
SAG1	1011	78	26/1011	26	AY187278, GQ253073, GQ253074, GQ253086-GQ253098, JX045355, JX045357-JX045359, JX045361-JX045367, JX045369-JX045397, JX045400-JX045421
locus 13 Chr. Ia	717	74	19/717	25	JX044860-JX044925, JX044928-JX044935
locus 6 Chr Ia	457	73	11/457	25	JX044553-JX044579, JX044581-JX044607, JX044609, JX044611-JX044628
MIC2 intron 1	412	75	9/412	22	JX044245, JX044247-JX044268, JX044270-JX044321
PK1	459	16	10/459	22	EU258505-EU258518, EU650327, EU650328
locus 8 Chr. Ia	747	74	15/747	21	JX044708-JX044717, JX044719-JX044782
locus 2 Chr Ia	731	73	14/731	20	JX044401-JX044456, JX044458, JX044460-JX044475
locus 14 Chr. Ia	742	73	14/742	19	JX044936, JX044939-JX044953, JX044955-JX045011
locus 26 Chr. Ia	744	74	14/744	19	JX045242-JX045266, JX045268-JX045277, JX045279-JX045317
locus 12 Chr. Ia	631	73	11/631	18	JX044783, JX044785, JX044787-JX044819, JX044821-JX044858

BY-NC-ND

TUB2 intron 1	447	77	6/447	14	AF249701-AF249703, AY143119, AY143121, AY143122, JX045496-JX045507, JX045509-JX045536, JX045538-JX045543, JX045545-JX045558, JX045561, JX045562, JX045564-JX045572
locus 7 Chr Ia	740	74	9/740	13	JX044629-JX044647, JX044649-JX044697, JX044699, JX044701-JX044705
Actin	1563	23	15/1563	10	JX045422-JX045443, JX045445
locus 19 Chr. Ia	746	74	4/746	6	JX045089, JX045090, JX045092-JX045126, JX045129-JX045165
TUB1 intron 1	501	6	2/501	4	AF249699, AF249700, AY143106, AY143107, AY143109, AY143110
locus 16 Chr.Ia	539	74	1/539	2	JX045013-JX045047, JX045050-JX045088

P: Polymorphic sites per kilobase
S: Number of polymorphic sites (including sites with gaps)
L: Length of sequences (with gaps)

BY-NC-ND

Annexe 2.3. S3 Table : Polymorphism and composition of the 13 *Toxoplasma gondii* populations.

Population	Number of strains	Number of SNPs					Haplotype diversity*	Hs †	Mean of hamming distances	Allelic richness‡
		GRA6	*GRA7*	*SAG3*	*UPRT1*	*UPRT7*				
Europe										
Austria	10	0	0	1	0	0	0.200	0.001	0.400	141
France	12	9	14	15	0	0	0.333	0.043	13.303	178
Portugal	12	8	14	15	0	0	0.250	0.131	40.090	177
Asia										
Turkey	13	6	9	15	9	2	0.272	0.050	15.636	181
China	14	10	20	10	12	8	0.357	0.090	27.164	200
Africa										
Ethiopia	11	8	14	15	0	0	0.230	0.067	20.615	177
Gabon	14	9	19	0	9	2	0.357	0.134	40.637	179
North America										
Minnesota	12	16	15	15	0	1	0.583	0.088	26.939	187
Pennsylvania	12	15	21	15	0	0	0.500	0.111	33.969	191
South America										
Guiana shield	17	19	15	18	18	20	1.000	0.133	39.588	231
Northeast Brazil	12	10	22	4	13	10	0.416	0.119	36.515	199
Southeast Brazil	15	13	22	2	18	10	0.666	0.166	49.866	205
Caribbean	14	14	24	18	7	2	0.785	0.128	38.703	205

Notes.
*: number of haplotypes / number of isolates
†: expected heterozygosity
‡: number of alleles

BY-NC-ND

Annexe 2.4. S4 Table : Position, size and primers of loci used in this study

Locus	Chromosome	Position of amplified fragment in chromosome	Size* (bp)	SNPs	Primers	Source
GRA6	X	7 193 682 - 7 194 472	607	26	F: GTAGCGTGCTTGTTGGCGAC R: TACAAGACATAGAGTGCCCC	[1]
GRA7	VIIa	2 604 391 - 2 605 143	677	30	F: TAAAAGGGATTACTGCGGCG R: CCCATATCCTACTGGCGGG	This study
SAG3	XII	511 192 - 511 871	638	27	F: CCAATGGTCGAAGGCAAGAG R: CTCCGGAAAACACAGTTCGG	This study
UPRT intron 1	XI	2 718 724 - 2 719 340	574	32	F: TCCAGGACATCATCACGAGG R: CCGCTTAGTGGCTTGATTCC	This study
UPRT intron 7	XI	2 722 126 - 2 722 885	657	25	F: TGGTCGTCGTCACTTGTTA R: GCAGCCTCACAACTAAAACT	[2]

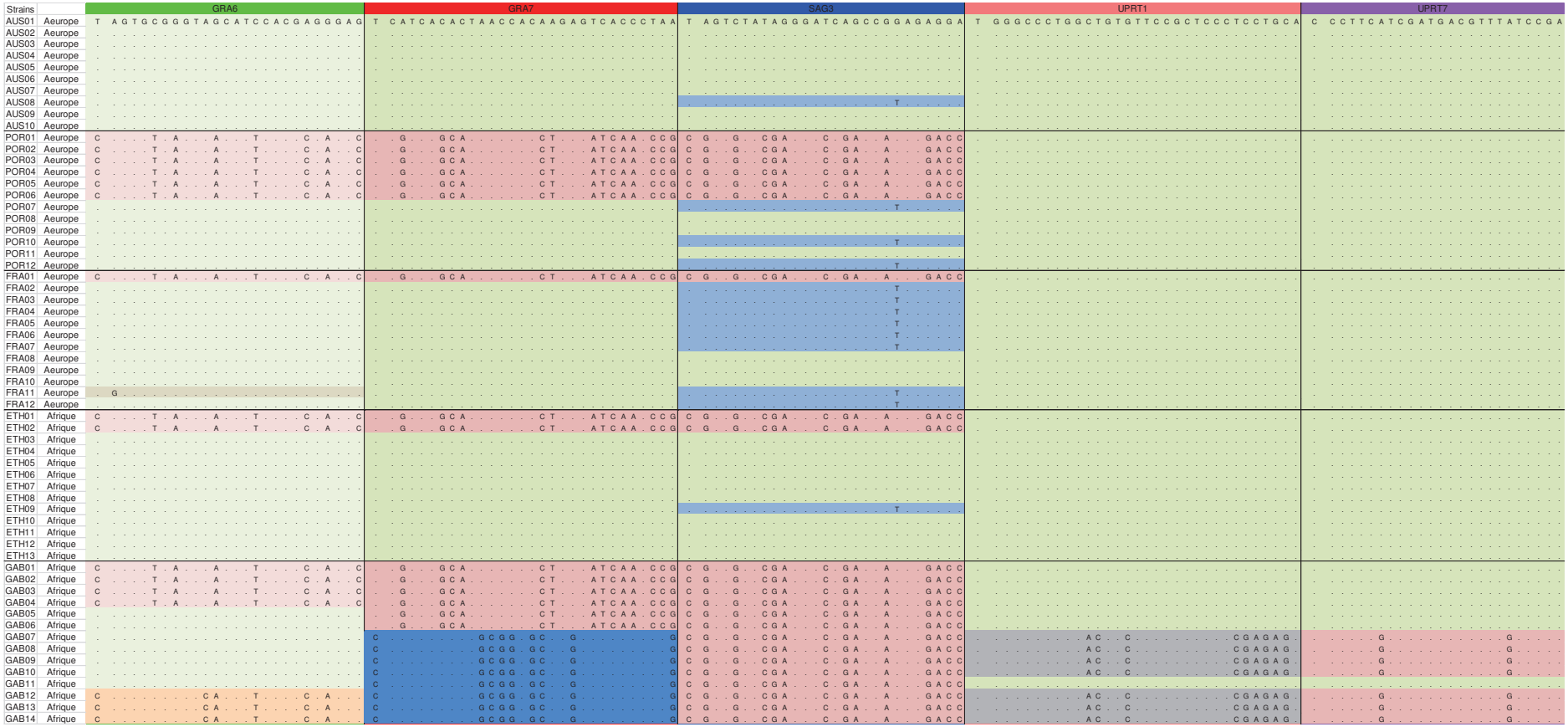
*: Size of fragments used in this study

1. Khan A, Fux B, Su C, Dubey JP, Darde ML, Ajioka JW, et al. Recent transcontinental sweep of *Toxoplasma gondii* driven by a single monomorphic chromosome. Proc Natl Acad Sci U S A. 2007;104: 14872–14877.

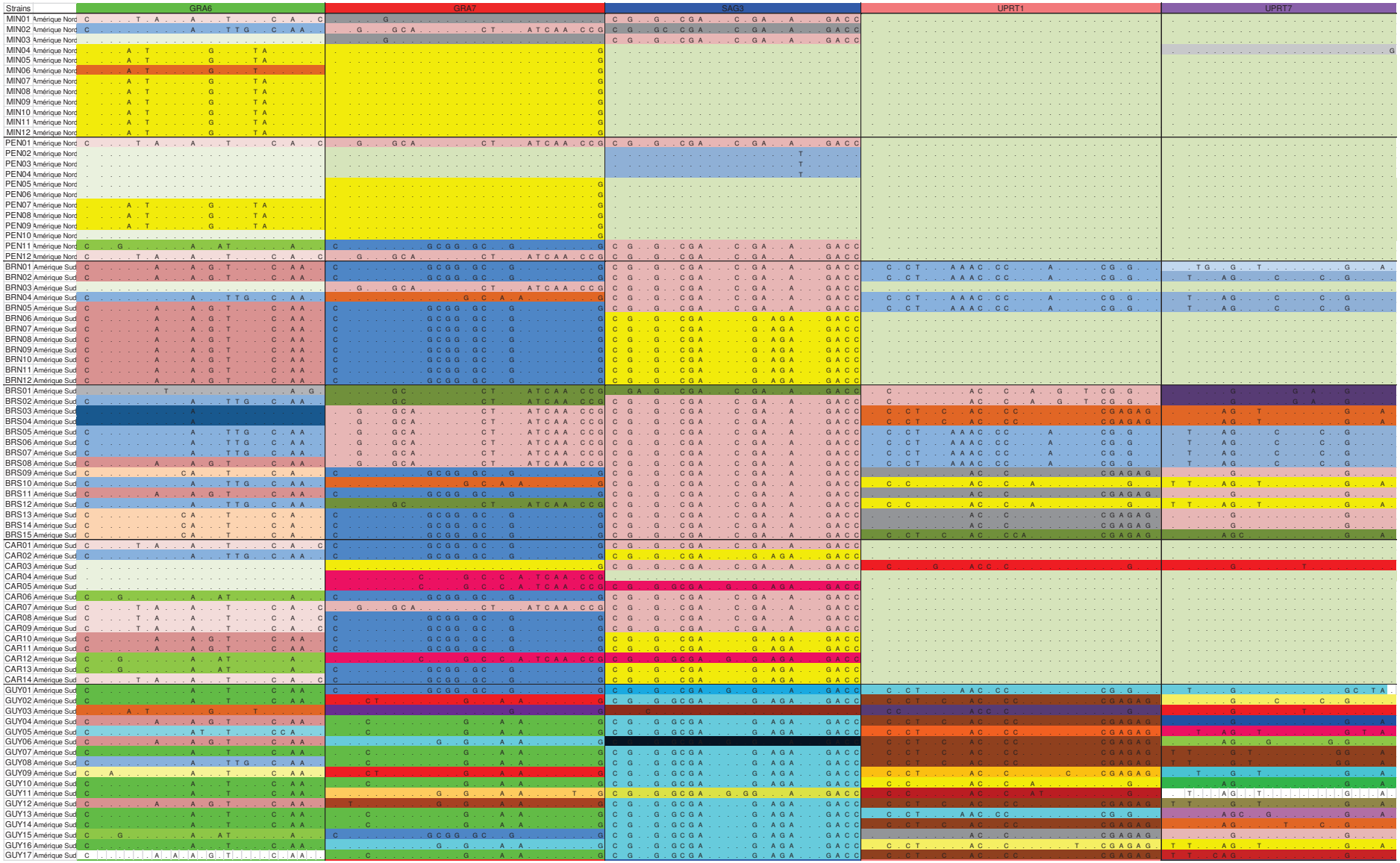
2. Fazaeli A, Carter PE, Darde ML, Pennington TH. Molecular typing of *Toxoplasma gondii* strains by GRA6 gene sequence analysis. Int J Parasitol. 2000;30: 637–642.

BY-NC-ND

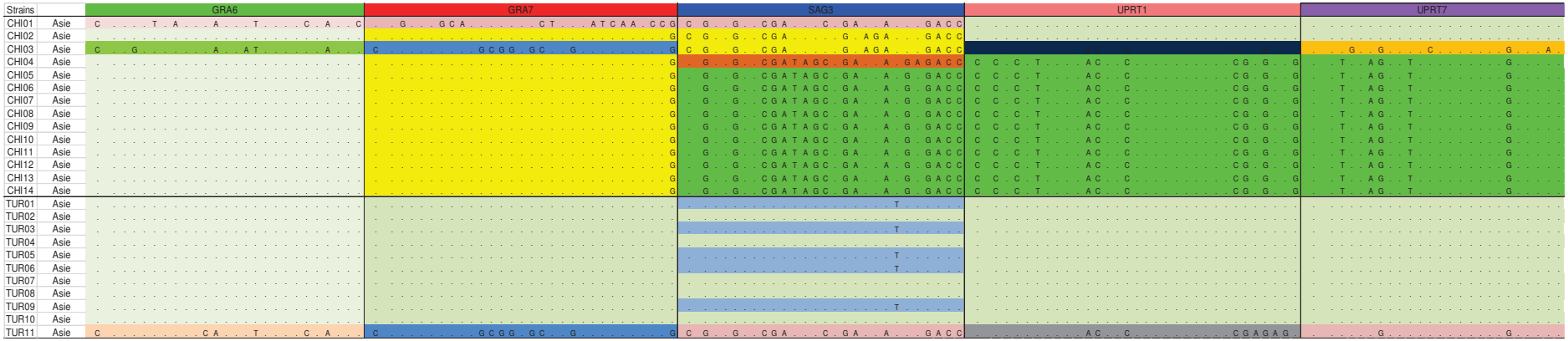
Annexe 2.5. S5 Table : List of haplotypes



BY-NC-ND



BY-NC-ND



This table does not take into account indels. Indels are:

Alignment *GRA6*:

Insertion GAT (268-270): GUY03, MIN04-MIN12, PEN07-PEN09

Alignment *GRA7:*

Deletion AAG (273-275): CAR03, CHI02, CHI04-CHI14, MIN04-MIN09, MIN11, MIN12, PEN05, PEN07-PEN09

Alignment *UPRT1*:

Insertion A (247): BRS01-BRS02, BRS10, BRS12, GUY10, GUY11

Alignment *UPRT7*:

Insertion TGTTC (463-467): GUY11

BY-NC-ND

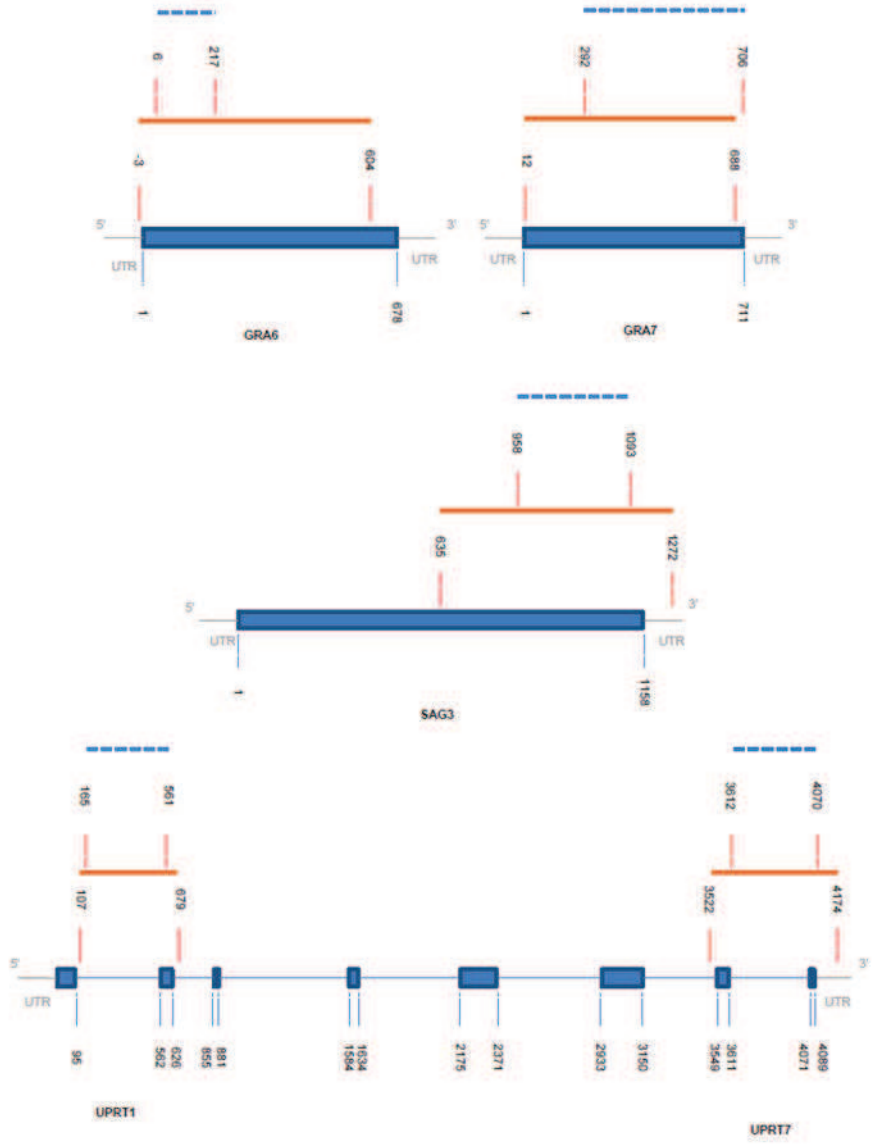
Annexe 2.6. S6 Table : Descriptive statistics of the data.

Gene	Exons			Introns		
	L	S	S’	L	S	S’
GRA6	604	102	29	3	0	0
GRA7	683	168	33	0	0	0
SAG3	527	63	21	114	21	6
UPRT1	65	2	0	513	110	33
UPRT7	82	5	2	603	115	28

L: sequence length. S: number of segregating sites (S’: without the outgroup).

BY-NC-ND

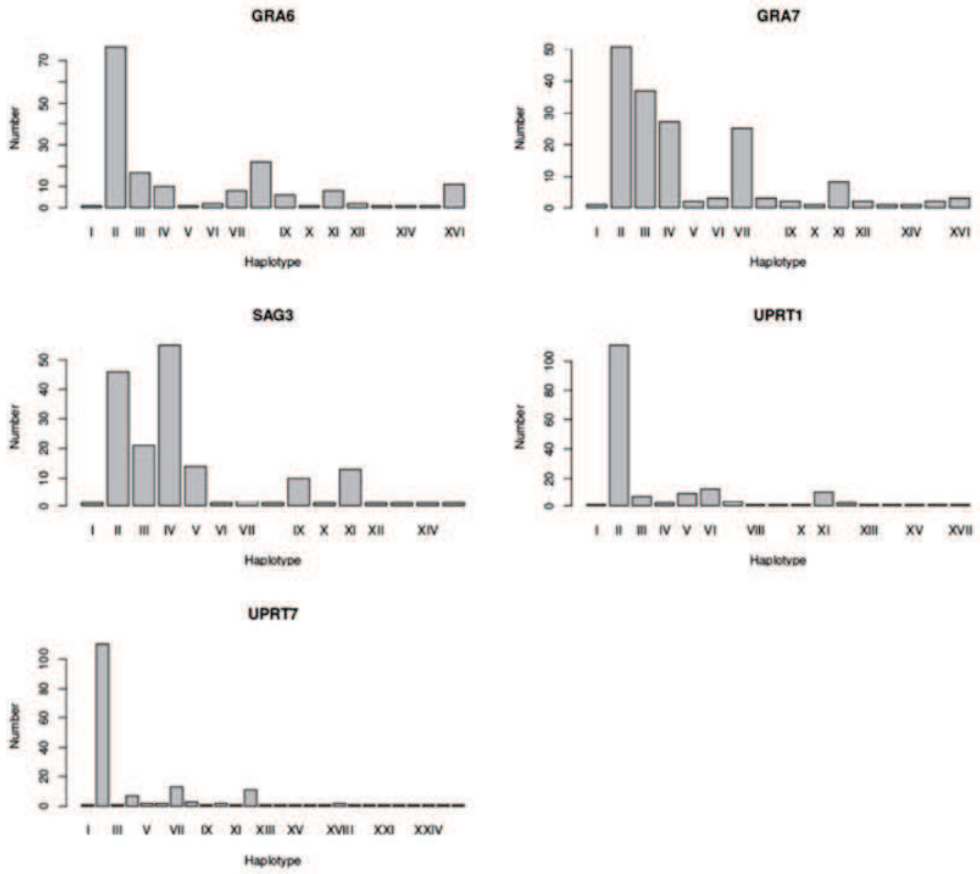
Annexe 2.7. S1 Figure



S1 Figure. Structure of *GRA6*, *GRA7, SAG3* and *UPRT* genes and location of amplified fragments in this study (continuous line), based on the reference strain ME49. For each gene, the locations of fragments retrieved from GenBank and used for selecting the markers are indicated (dotted line).

BY-NC-ND

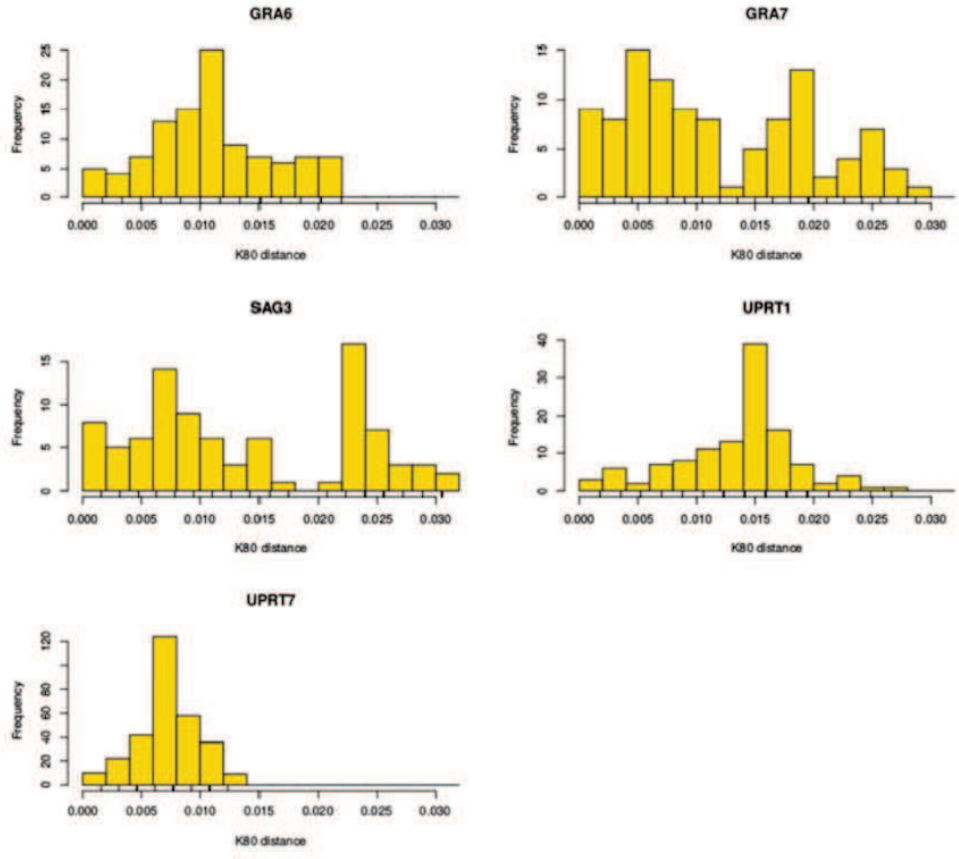
Annexe 2.8. S2 Figure



S2 Figure. Distribution of haplotypes frequencies.

BY-NC-ND

Annexe 2.9. S3 Figure



S3 Figure. Distribution of pairwise K80 distances.

BY-NC-ND

Table des illustrations

Figure 1 : Schéma de la structure d'un tachyzoïte de *Toxoplasma gondii.* 7

Figure 2 : Arbre phylogénétique (Neighbour joining) illustrant les 4 grands groupes d'Apicomplexes. 8

Figure 3 : Cycle de vie de *Toxoplasma gondii* 12

Figure 4 : Cycle de vie d'*Hammondia hammondi* 12

Figure 5 : Distribution géographique des génotypes de *Toxoplasma gondii.* 15

Figure 6 : Structure de population de *Toxoplasma gondii.* 16

Figure 7 : Arbre phylogénétique des félidés, reconstruit avec la méthode de maximum de vraisemblance et à partir de 18853 bp d'ADN nucléaire. 23

Figure 8 : Les migrations des Félidés expliquant la mise en place des huit lignées. 25

Figure 9 : Migrations des différentes espèces à l'intérieur des lignées, expliquant leur distribution actuelle. 26

Figure 10 : Représentation schématique des 3 modèles d'évolution proposés pour *Toxoplasma gondii* 28

Figure 11 : Modèle proposé pour la création des 3 types clonaux de *Toxoplasma gondii.* 29

Figure 12 : Modèle pour la génération des 11 premiers haplogroupes à partir de groupes anciens ressemblant aux haplogroupes 2, 4, 6 et 9. 30

Figure 13 : Modèle proposé pour l'expansion des populations clonales de *Toxoplasma gondii* en Amérique du Nord et en Europe 31

Figure 14 : Distribution géographique des populations et nombre de souches par population. 38

Figure 15 : Structure des gènes *GRA6*, *GRA7*, *SAG3*, *UPRT1* et *UPRT7* et localisation des fragments amplifiés dans cette étude (ligne continue orange), basée sur la souche de référence ME49. 47

Figure 16 : Nombre et fréquences des haplotypes pour *GRA6*, *GRA7*, *SAG3*, *UPRT1* et *UPRT7*. 54

Figure 17 : Distributions des distances K80 deux à deux pour *GRA6, GRA7, SAG3, UPRT1* et *UPRT7* 55

BY-NC-ND

Figure 18 : Nombre et fréquences des haplotypes présents en Asie, dans les Caraïbes, en Europe, en Afrique, en Amérique du Nord et en Amérique du Sud. 56

Figure 19 : Principe de l'agglomération du Neighbor-Joining. L'algorithme commence avec un arbre en étoile et agglomère une paire de taxons. 59

Figure 20 : Composition de quelques modèles d'évolution de séquences nucléotidiques. ...61

Figure 21 : Arbre NJ enraciné avec *Hammondia hammondi*. 64

Figure 22 : Arbre BioNJ enraciné avec *Hammondia hammondi*. 65

Figure 23 : Arbre en FastMe enraciné avec *Hammondia hammondi*. 66

Figure 24 : Arbre Maximum de Vraisemblance sur les régions codantes GRA6, GRA7 et SAG3, enraciné avec *Hammondia hammondi*. 68

Figure 25 : Arbre Maximum de vraisemblance sur les régions non-codantes de GRA6, GRA7 et SAG3, enraciné avec *Hammondia hammondi*. 68

Figure 26 : Arbre Maximum de Vraisemblance sur les régions codantes d'*UPRT1* et *7*, enraciné avec *Hammondia hammondi* 69

Figure 27 : Arbre Maximum de Vraisemblance sur les régions non-codantes d'*UPRT1* et *7*, enraciné avec *Hammondia hammondi*. 69

Figure 28 : Consensus des 4 arbres correspondants aux partitions suivantes régions codantes/ régions non-codantes, *GRA6*, *GRA7*, *SAG3* / *UPRT1*, *UPRT7* 70

Figure 29 : Les migrations successives conduisent à une perte progressive de la diversité de la population. 72

Figure 30 : Représentations des premières étapes de l'analyse spatiale. 74

Figure 31 : Répartition historique de *Toxoplasma gondii* inférée à partir des 5 marqueurs. ..75

Figure 32. Principe de la datation moléculaire 77

Figure 33 : Processus de coalescence……78

BY-NC-ND

Table des tableaux

Tableau 1 : Polymorphisme des différentes régions d'ADN de *Toxoplasma gondii* présentes dans GenBank.....42

Tableau 2 : Position, longueur et amorces des loci utilisés dans cette étude. 45

Tableau 3 : Polymorphisme et composition des 13 populations……..……..……52

BY-NC-ND

Table des matières

Introduction....5

Chapitre I. Classification, biologie et histoire évolutive de *Toxoplasma gondii*....6

I.1. Place de *Toxoplasma gondii* parmi les Apicomplexes....6

I.1.1. Les Apicomplexes....6

I.1.2. Classification des Apicomplexes....7

I.1.3. Relations de parenté dans la sous-famille des Toxoplasmatinae....9

I.2. Cycle de vie de *T. gondii* et diversité génétique....10

I.2.1. Cycle de vie....10

I.2.1.1. Cycle de vie de *T. gondii* et comparaison avec *H. hammondi*....10

I.2.2. Diversité génétique de *T. gondii*....13

I.2.2.1. Marqueurs utilisés pour génotyper *T. gondii*....13

I.2.2.2. Principaux résultats....14

I.2.2.3. Les moteurs de la diversité génétique de *T. gondii*....17

I.2.3. Séroprévalence et formes cliniques de la toxoplasmose avec un focus sur l'Amérique du Sud....18

I.3. Etat de l'art sur l'histoire évolutive de *T. gondii* et de ses hôtes définitifs....22

I.3.1. Histoire évolutive des félidés....23

I.3.1.1. La phylogénie....23

I.3.1.2. Les migrations....25

I.3.1.3. L'échange des faunes entre Amérique du Nord et Amérique du Sud....26

I.3.2. Origine des 3 types clonaux de *T. gondii*....27

I.3.3. Débat sur l'origine géographique de *T. gondii*....32

I.4. Objectifs de la thèse....33

Chapitre II. Etude expérimentale....37

II.1. Le jeu de données....37

II.1.1. Sélection des souches....37

II.1.2. Sélection des marqueurs....38

II.1.3. Amplification et séquençage....45

II.2. Diversité génétique de *T. gondii*....48

II.2.1. Estimer la diversité génétique....48

II.2.2. Calcul de la diversité génétique des populations de *T. gondii*....50

II.2.2.1. Détermination du polymorphisme par marqueur....50

II.2.2.2. Détermination de la diversité génétique intra-populationnelle....50

II.2.3. Résultats....51

II.2.3.1. Les sites polymorphes au sein des marqueurs....51

II.2.3.2. Les diversités génétiques des populations....55

II.2.4. Discussion....57

II.2.4.1. Polymorphisme et nature des marqueurs....57

II.2.4.2. Diversité intra-populationnelle et divergence entre les indices de diversité génétique....57

II.3. Phylogénie....58

BY-NC-ND

II.3.1. Les phylogénies moléculaires 58
II.3.1.1. Les méthodes basées sur les mesures de distances 58
II.3.1.2. Les méthodes basées sur les caractères discrets 60
II.3.2. Phylogénie de *T. gondii* 62
II.3.2.1. Choix du groupe externe 62
II.3.2.2. Analyses exploratoires : méthodes de distance 62
II.3.2.3. Phylogénie : méthode de Maximum de vraisemblance 62
II.3.2.4. Consensus des arbres 63
II.3.3. Résultats 63
II.3.3.1. Les arbres de distances 63
II.3.3.2. Les arbres en maximum de vraisemblance 67
II.3.3.3. Le consensus 70
II.3.4. Discussion 70
II.4. Analyse spatiale 72

II.4.1. Principe de la relation diversité génétique/distance géographique 72
II.4.2. Détermination de l'origine géographique de *Plasmodium falciparum* 72
II.4.3. Application au cas de *Toxoplasma gondii* 73
II.4.4. Résultats 75
II.4.5. Discussion 76
II.5. Datations 76

II.5.1. Objectif et méthodes de datation 76
II.5.1.1. Datation moléculaire 77
II.5.1.2. Méthode de coalescence 78
II.5.2. Datation du TMRCA des lignées actuelles de *T. gondii* 79
II.5.2.1. Calcul du taux de mutation 79
II.5.2.2. Datation moléculaire 79
II.5.2.3. Méthode de coalescence 79
II.5.3. Résultats 80
II.5.3.1. Taux de mutation 80
II.5.3.2. Les datations du TMRCA 80
II.5.4. Discussion 80
II.6. Discussion générale 81

Perspectives et conclusion 84
Références bibliographiques 85
Annexes 95

BY-NC-ND

Sur l'origine de *Toxoplasma gondii* : approches phylogénétique et spatialement-explicite pour la détermination de l'origine géographique d'un parasite ubiquiste

Toxoplasma gondii, protozoaire ubiquitaire chez les mammifères et les oiseaux, est l'agent étiologique de la toxoplasmose, une maladie posant un réel problème de santé publique dans le monde avec environ 200 000 nouveaux cas de toxoplasmose congénitale chaque année. Il a été montré que la sévérité clinique de la toxoplasmose variait en fonction des régions géographiques, avec en particulier l'Amérique du Sud qui paie le plus lourd tribu de cette maladie. Malheureusement, les mécanismes de ces disparités géographiques sont encore peu compris et l'origine géographique ainsi que l'histoire évolutive du pathogène sont encore incertaines. Une collection mondiale de 168 isolats de *T. gondii* recueillis dans 13 populations de 5 continents a été séquencée pour cinq fragments de gènes (140 single nucleotide polymorphisms à partir de 3153 bp par isolat). La phylogénie basée sur les méthodes de Maximum de vraisemblance avec une estimation de l'âge du plus récent ancêtre commun (TMRCA) et des analyses géostatistiques ont été réalisées afin d'inférer l'origine hypothétique de *T. gondii*. Nous montrons que les souches actuelles de ce parasite ont vraisemblablement évolué à partir d'un ancêtre Sud-Américain il y a environ 1,5 million d'années et avons reconstruit la propagation mondiale du pathogène qui a suivi. Cette émergence est beaucoup plus récente que l'apparition de la forme ancestrale de *T. gondii* il y a environ 11 Ma et est postérieure à l'arrivée des félidés dans cette partie du monde. Nous proposons que la lignée ancestrale de *T. gondii* ait été introduite en Amérique du Sud avec les félidés et que l'évolution de l'infectivité orale des kystes tissulaires à travers le carnivorisme ainsi que la diversification des félidés dans cette région du monde a permis l'apparition d'une nouvelle souche ayant une capacité de transmission beaucoup plus efficace que la lignée ancestrale, ce qui lui a permis de la supplanter et d'avoir une distribution pandémique.

Mots-clés : *Toxoplasma gondii*, phylogéographie, phylogénie, date du plus récent ancêtre commun, diversité génétique

On the origin of *Toxoplasma gondii*: phylogenetic and spatially explicit approaches for the identification of the geographical origin of an ubiquitous parasite

Toxoplasma gondii, a protozoan found ubiquitously in mammals and birds, is the etiologic agent of toxoplasmosis, a disease causing substantial Public Health burden worldwide, including about 200,000 new cases of congenital toxoplasmosis each year. Clinical severity has been shown to vary across geographical regions with South America exhibiting the highest burden. Unfortunately, the drivers of these heterogeneities are still poorly understood, and the geographical origin and historical spread of the pathogen worldwide are currently uncertain. A worldwide sample of 168 *T. gondii* isolates gathered in 13 populations was sequenced for five fragments of genes (140 single nucleotide polymorphisms from 3,153 bp per isolate). Phylogeny based on Maximum likelihood methods with estimation of the time to the most recent common ancestor (TMRCA) and geostatistical analyses were performed for inferring the putative origin of *T. gondii*. We show that extant strains of the pathogen likely evolved from a South American ancestor, around 1.5 million years ago, and reconstruct the subsequent spread of the pathogen worldwide. This emergence is much more recent than the appearance of ancestral *T. gondii*, believed to have taken place about 11 My ago, and follows the arrival of felids in this part of the world. We posit that an ancestral lineage of *T. gondii* likely arrived in South America with felids and that the evolution of oral infectivity through carnivorism and the



radiation of felids in this region enabled a new strain to outcompete the ancestral lineage and undergo a pandemic radiation.

Keywords: *Toxoplasma gondii*, phylogeography, phylogeny, time to the most recent common ancestor (TMRCA), genetic diversity

