

Sequenciação de nova geração: o paradigma dos parasitas

New generation sequencing: the parasite paradigm

Anabela Vilares¹, Vítor Borges², Daniel Sampaio³, Luís Vieira³, Idalina Ferreira¹, Susana Martins¹, Tânia Reis¹, João Paulo Gomes², Maria João Gargaté¹

m.joao.gargate@insa.min-saude.pt

(1) Laboratório Nacional de Referência de Infecções Parasitárias e Fúngicas. Departamento de Doenças Infecciosas, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal.

(2) Unidade de Bioinformática. Departamento de Doenças Infecciosas, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal.

(3) Unidade de Tecnologia e Inovação. Departamento de Doenças Infecciosas, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal.

_Resumo

Toxoplasma gondii é um protozoário intracelular obrigatório responsável pela Toxoplasmose, infetando diferentes espécies incluindo o Homem. Na Europa e no Norte da América, *T. gondii* tem uma população clonal constituída por tipo I, II, III. No entanto, estudos recentes têm descrito uma maior diversidade genética, com a identificação de estirpes recombinantes e atípicas mais virulentas. Em Portugal, o conhecimento dos genótipos circulantes de *T. gondii* é limitado. Neste sentido, estudou-se a diversidade genética de 68 estirpes, incluindo 51 estirpes de *T. gondii* isoladas em Portugal, a estirpe de referência RH e 16 outras estirpes de referência. Neste estudo realizou-se a genotipagem clássica (Sag2), PCR multiplex de 5 microssatélites e uma combinação de sequenciações por Sanger e NGS de oito loci responsáveis pela virulência. A genotipagem clássica (tipo I, II, III) foi obtida em 100% das estirpes e mostrou que a maioria das estirpes isoladas em Portugal eram do tipo II (70,6%; 36/51) e as restantes 15 do tipo I. A análise de microssatélites por PCR multiplex aumentou o poder discriminatório em apenas mais um tipo identificado. Contudo, quando se adicionaram as metodologias de sequenciação por Sanger e NGS o poder discriminatório aumentou para 36 tipos diferentes. A combinação destas metodologias (Sanger/NGS) permitiram a identificação de uma estrutura mosaico nos isolados de *T. gondii* em Portugal, não conseguida anteriormente por qualquer uma das outras tecnologias utilizadas.

_Abstract

Toxoplasma gondii is an obligate intracellular protozoan parasite, which is responsible for toxoplasmosis in different species, including humans. *T. gondii* have a distinct clonal population structure composed of type I, II and III lineages in North America and Europe. However, recent studies demonstrated a higher diversity given by recombinant and atypic strains, which are believed to be more virulent. The scenario of the distribution of *T. gondii* genotypes is lacking in Portugal. In order to achieve this knowledge we studied the genetic diversity of 68 *T. gondii* strains, included 51 isolated in Portugal, a reference strain RH and 16 others reference strains). We performed classical genotyping (Sag2), 5 microsatellites multiplex PCR and a combination of Sanger sequencing and Next Generation Sequencing (NGS) of eight loci likely responsible for virulence. The classical genotyping was achieved in 100% of strains. The majority of strains (70.6%; 36/51) isolated in Portugal were type II whereas the others 15 were type I. However, when Sanger and NGS sequencing methodologies were added, discriminatory power increased to 36 different

types. The combination of these methodologies (Sanger/NGS) allowed the identification of a mosaicism in the isolates of *T. gondii* in Portugal, previously not achieved by any of the other technologies used.

_Introdução

Toxoplasma gondii é o parasita responsável pela Toxoplasmose infetando diferentes hospedeiros incluindo os humanos (1,2). A infeção por *T. gondii* é na sua maioria assintomática no indivíduo imunocompetente, sendo extremamente agressiva no recém-nascido, causando cegueira e/ou doença com sequelas do sistema nervoso central. Nos indivíduos imunodeprimidos a doença pode levar à morte (3).

Nos últimos anos diferentes investigadores têm adotado *Next Generation Sequencing* (NGS) em detrimento da sequenciação clássica (Sanger sequencing) (4,5). Apesar da sequenciação por Sanger permitir a identificação de alelos com uma frequência superior a 50%, quando existe mistura de genótipos, recombinação ou espécies com alelos de frequência inferior, estes são "mascarados", perdendo-se assim a sua identificação (6,7).

Existem três tipos de NGS para sequenciação de DNA: 1) a sequenciação do genoma completo (WGS, na sigla em inglês), 2) a sequenciação de todos os exões (WES) e 3) a sequenciação de zonas alvo do genoma. WGS avalia todo o conteúdo do genoma de um indivíduo em estudo, enquanto que WES sequencia apenas zonas do genoma codificantes de proteínas. Em contrapartida a sequenciação de zonas alvo do genoma, utiliza regiões específicas do genoma (intrões ou/e exões), que têm

como base alguns *locus* específicos ligados a mecanismos de patogenicidade ou resistência conhecidos (8).

No caso dos parasitas, como *Toxoplasma gondii* que tem um genoma com 14 cromossomas e cerca de 65 Mb, WGS requer tecnologias com equipamentos dispendiosos, levando a maioria dos laboratórios a utilizarem apenas zonas específicas do genoma do protozoário (9-12).

Alguns estudos de filogenia e resistência a fármacos antiparasitários têm sido publicados utilizando a tecnologia NGS (9,13), contudo não há qualquer estudo deste género aplicado a estirpes de *Toxoplasma gondii* isoladas em Portugal.

_Objetivo

Este estudo tem como objetivo demonstrar a mais-valia das novas metodologias para a caracterização genética de estirpes de *T. gondii* de origem humana isoladas no Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA) entre 1994 e 2013.

_Material e métodos

Amostra

As amostras utilizadas neste estudo pertencem à coleção de estirpes de *Toxoplasma gondii* do INSA, tendo sido isoladas em humanos entre 1994 e 2013.

Entre 1994 e 2013, foram isoladas 51 estirpes de *T. gondii*, 50 associados à toxoplasmose congénita (CT) e uma proveniente de uma biópsia cerebral (BC) pertencente a um doente com SIDA. Foram ainda estudadas 17 estirpes internacionais de controlo.

PCR, Sanger sequencing and NGS

Foram efetuadas PCR das duas sequências terminais 5' e 3' do gene *Sag2* (genotipagem clássica em I, II, III com base na RFLF), PCR multiplex de 5 microssatélites e PCR em oito zonas alvo (*Sag1*, *Sag2*- sequencias terminais 5' e 3', *Sag3*, CB21-4, M102, L363, PK1, *Gra6*) para proceder à sequenciação (Sanger e NGS) em todos os isolados. Para aplicação da

metodologia de NGS os produtos de PCR de cada isolado foram purificados antes de se proceder à preparação da biblioteca genómica e subsequentemente sequenciados no equipamento MiSeq (Illumina).

_Resultados

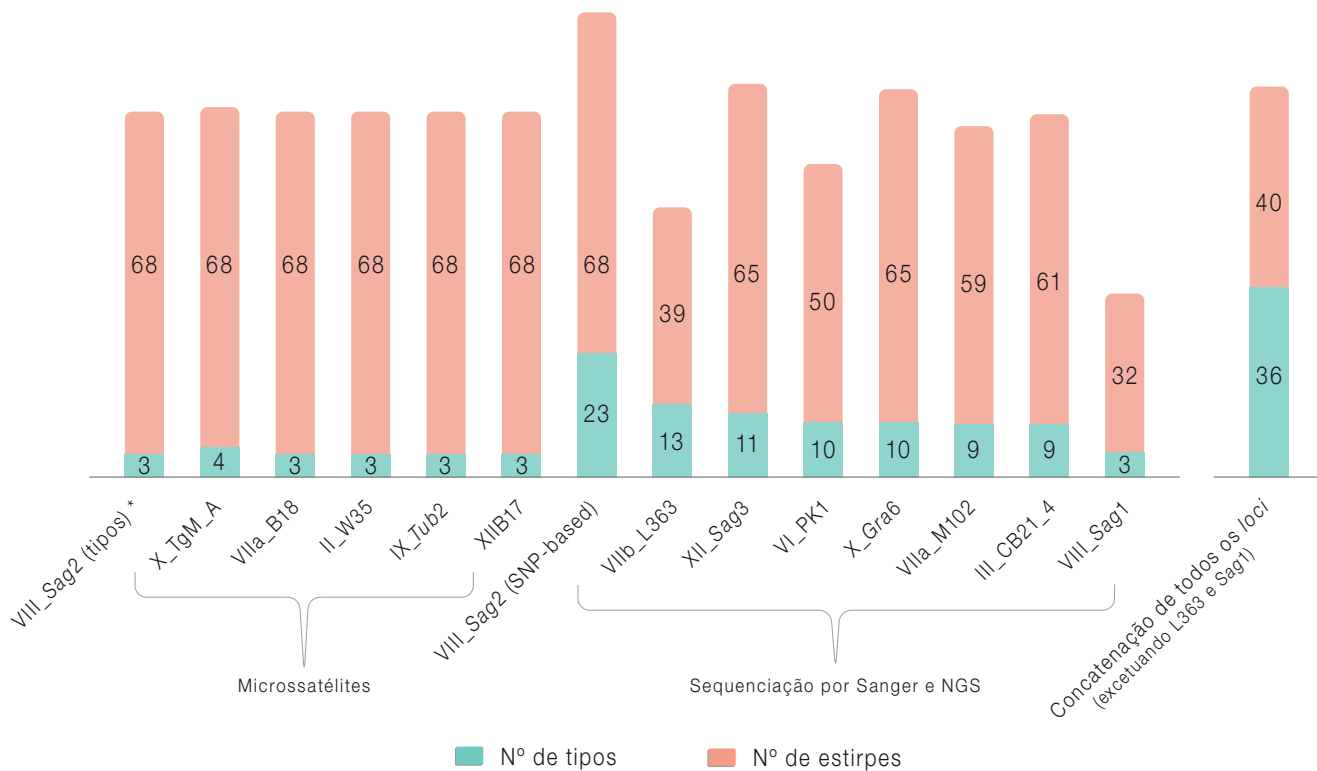
A genotipagem clássica pelo gene *Sag2* (I,II,III) foi conseguida em 100% das amostras, sendo a maioria dos isolados de *Toxoplasma gondii* pertencentes ao tipo II (70,6%; n=36) e os restantes do tipo I (29,4%, n=15) nos isolados de humanos em Portugal. Nestes mesmos isolados não se identificou qualquer estirpe do tipo III.

Através da análise conjunta dos isolados de Portugal e das estirpes de referência foi possível verificar que a introdução da PCR multiplex dos 5 microssatélites não aumentou o poder discriminatório por *locus* (máximo de 4 tipos).

Após a inclusão dos loci para os quais se determinou a sequência nucleotídica (por Sanger e/ou NGS) o poder discriminatório aumentou consideravelmente. De facto, 6 dos 8 *loci* permitiram discriminar mais de 10 tipos, sendo que *Sag2* permitiu identificar 23 tipos entre as 68 estirpes analisadas. É, ainda, de destacar, que em 40 estirpes foi possível obter dados para todos os *loci*, com a exceção do L363 e *Sag1* em que se obteve baixo sucesso de sequenciação. A combinação dos perfis obtidos permitiu diferenciar 36 tipos das 40 estirpes estudadas (gráfico 1).

Para além deste elevado poder discriminatório, esta nova abordagem desvendou perfis recombinantes reveladores de estruturas genéticas mosaico em *T. gondii* (estudo em progresso).

Gráfico 1: Comparação de perfis para os diferentes loci das estirpes de *Toxoplasma gondii* analisadas.



* Sag2 - Classificação clássica (I, II, III) com base na metodologia de RFLP

Discussão e conclusão

Atualmente a metodologia NGS está presente e é indispensável na investigação genómica. A utilização dos dados obtidos por NGS permite a identificação de marcadores em larga escala do genoma completo, bem como a identificação de eventos de recombinação (14). O presente estudo, que constitui a primeira abordagem em Portugal utilizando a técnica NGS aplicada a *T. gondii*, permitiu obter um elevado poder discriminatório relativamente às outras metodologias anteriormente utilizadas, principalmente no que se refere aos fenómenos de recombinação neste parasita (gráfico 1). À semelhança dos avanços da epidemiologia molecular de outras doenças infecciosas estas metodologias permitem uma melhor compreensão do perfil genético das estirpes circulantes em Portugal, com ganho para a vigilância da infeção congénita, para a

possível deteção de surtos (embora raros) e subsequente implementação de medidas de controlo.

Apesar de atualmente a terapêutica utilizada ser independente do genótipo identificado pelo método clássico, não é de descartar que com a revelação de novas relações de fenótipo/genótipo identificadas por esta nova tecnologia, este facto poderá contribuir no futuro para uma melhor adequação da terapêutica quer na infeção congénita, quer nas infeções adquiridas pelos imunodeprimidos.

Agradecimentos

Este trabalho foi financiado pela Fundação para a Ciência e Tecnologia (SFRH/BD/75154/2010). Agradece-se à Doutora Cristina Furtado pela revisão do artigo.

Referências bibliográficas:

- (1) Su C, Evans D, Cole RH, et al. Recent expansion of *Toxoplasma* through enhanced oral transmission. *Science*. 2003;299(5605):414-6. <http://science.sciencemag.org/content/299/5605/414.long>
- (2) Dubey JP, Jones JL. *Toxoplasma gondii* infection in humans and animals in the United States. *Int J Parasitol*. 2008;38(11):1257-78.
- (3) Remington G. Schizophrenia, antipsychotics, and the metabolic syndrome: is there a silver lining? *Am J Psychiatry*. 2006;163(7):1132-4.
- (4) Cooper C, Keatley S, Northover A, et al. Next generation sequencing reveals widespread trypanosome diversity and polyparasitism in marsupials from Western Australia. *Int J Parasitol Parasites Wildl*. 2018;7(1):58-67. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6031965/>
- (5) Wilkinson MJ, Szabo C, Ford CS, et al. Replacing Sanger with Next Generation Sequencing to improve coverage and quality of reference DNA barcodes for plants. *Sci Rep*. 2017;7:46040. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5388885/>
- (6) Talundzic E, Plucinski MM, Biliya S, et al. Advanced Molecular Detection of Malarone Resistance. *Antimicrob Agents Chemother*. 2016;60(6):3821-3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4879379/>
- (7) Paparini A, Yang R, Chen L, et al. Cryptosporidium in fish: alternative sequencing approaches and analyses at multiple loci to resolve mixed infections. *Parasitology*. 2017;144(13):1811-1820.
- (8) McCombie WR, McPherson JD, Mardis ER. Next-Generation Sequencing Technologies. *Cold Spring Harb Perspect Med*. 2018 Nov 26. pii: a036798. [Epub ahead of print]
- (9) Lalremruata A, Jeyaraj S, Engleitner T, et al. Species and genotype diversity of *Plasmodium* in malaria patients from Gabon analysed by next generation sequencing. *Malar J*. 2017;16(1):398. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5627438/>
- (10) Nag S, Dalgaard MD, Kofoed PE, et al. High throughput resistance profiling of *Plasmodium falciparum* infections based on custom dual indexing and Illumina next generation sequencing-technology. *Sci Rep*. 2017;7(1):2398. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5445084/>
- (11) Rao PN, Uplekar S, Kayal S, et al. A Method for Amplicon Deep Sequencing of Drug Resistance Genes in *Plasmodium falciparum* Clinical Isolates from India. *J Clin Microbiol*. 2016;54(6):1500-11. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4879288/>
- (12) Talundzic E, Ndiaye YD, Deme AB, et al. Molecular Epidemiology of *Plasmodium falciparum* kelch13 Mutations in Senegal Determined by Using Targeted Amplicon Deep Sequencing. *Antimicrob Agents Chemother*. 2017;61(3). pii: e02116-16. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5328579/>
- (13) Talundzic E, Ravishankar S, Kelley J, et al. Next-Generation Sequencing and Bioinformatics Protocol for Malaria Drug Resistance Marker Surveillance. *Antimicrob Agents Chemother*. 2018;62(4). pii: e02474-17. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5913988/>
- (14) Ajzenberg D, Bañuls AL, Su C, et al. Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. *Int J Parasitol*. 2004;34(10):1185-96.