

### III Workshop de Dissertações de Mestrado do PPgSI (2016)

#### FICHA DA PESQUISA<sup>i</sup>

DADOS GERAIS					
<b>Título do projeto de pesquisa</b>	Uma técnica de comparação de genomas bacterianos para identificação dos genes mais representativos para a classificação filogenética				
<b>Orientando</b>	Vivian Mayumi Yamasaki Pereira				
<b>Orientador(es)</b>	Luciano Antonio Digiampietri				
<b>Momento atual</b>	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input checked="" type="checkbox"/> 4º semestre	<input type="checkbox"/> 5º semestre	<input type="checkbox"/> 6º semestre
<b>Qualificação</b>	<input checked="" type="checkbox"/> Qualificação já realizada em: 29/4/2016			<input type="checkbox"/> Qualificação planejada para: ___/___/___	
<b>Defesa</b>	Prazo máximo para depósito: 26/2/2018			Depósito planejado para: 1/2/2017	
<b>Linha e Área de pesquisa</b>	<input type="checkbox"/> Gestão e desenvolvimento de Sistemas: <input type="checkbox"/> BD <input type="checkbox"/> Gestão de SI <input type="checkbox"/> Eng. de Software <input type="checkbox"/> IHC		<input checked="" type="checkbox"/> Inteligência de Sistemas: <input checked="" type="checkbox"/> IA <input type="checkbox"/> Rec. de Padrões <input type="checkbox"/> Proc. Gráfico		
<b>Área de aplicação</b>	<input type="checkbox"/> Ambientes Corporativos <input checked="" type="checkbox"/> Bioinformática <input type="checkbox"/> Biometria <input type="checkbox"/> Economia	<input type="checkbox"/> Educação <input type="checkbox"/> Educação a Distância <input type="checkbox"/> Internet <input type="checkbox"/> Jogos	<input type="checkbox"/> Linguagem Natural <input type="checkbox"/> Linguística <input type="checkbox"/> Processos de Negócio <input type="checkbox"/> Química	<input type="checkbox"/> Redes Sociais <input type="checkbox"/> Robótica <input type="checkbox"/> Saúde <input type="checkbox"/> [outro – escrever]	
<b>Publicações</b>	Até o presente momento, há três trabalhos já publicadas. O artigo completo “Um sistema para análise de genomas a partir de metagenomas” foi publicado em um workshop (Brazilian e-Science Workshop 2015). Já os pôsteres “Bacterial whole genome comparison: a systematic literature review” e “A systematic review of phylogenetic analysis of a specific gene” foram publicados em uma conferência (X-Meeting 2015).				

#### DESCRIÇÃO DO PROJETO DE PESQUISA

<b>Contextualização / motivação</b>	A comparação entre genomas é importante para a classificação filogenética e para a identificação fenômenos biológicos que ocorrem nos organismos que, por sua vez, pode auxiliar a descoberta de causas de doenças.
<b>Problema de pesquisa</b>	O problema de pesquisa trata-se de verificar se há características que permitem classificar os organismos (inclusive os mais próximos evolutivamente) de acordo com sua real história evolutiva, tal que seja possível identificar a filogenia de maneira mais precisa.
<b>Objetivo geral da pesquisa</b>	O objetivo geral deste projeto é propor e implementar ferramentas computacionais para realizar a comparação entre genomas de bactérias, levando-se em consideração tanto a sequência de nucleotídeos quanto de genes, visando identificar os genes mais representativos para a classificação filogenética.
<b>Trabalhos relacionados</b>	Khripet (2005) realizou comparações proteômicas que resultaram em uma nova medida de distância, levando-se em conta a ordem e o conteúdo gênico, para a reconstrução filogenética. Snel, Bork e Huynen (1999) também utilizaram o conteúdo gênico (baseado na presença ou não de possíveis pares de genes ortólogos) para calcular as distâncias entre os genomas e gerar suas filogenias, que foram condizentes com as geradas com o gene 16S rRNA. Por fim, House e Fitz-Gibbon (2001) utilizaram a presença ou não de famílias de genes homólogos nos genomas para gerar as filogenias, que também eram similares às geradas a partir do gene 16S rRNA.
<b>Justificativa e relevância</b>	Apesar da importância da comparação de genomas e da análise filogenética, não foram encontradas metodologias robustas e detalhadas que realizam uma comparação completa, considerando diversos genes de genomas de bactérias. A metodologia mais utilizada é a comparação utilizando o gene 16S rRNA que, para genomas muito próximos evolutivamente, pode não refletir a filogenia real de bactérias muito próximas. Além disso, abordagem proposta neste projeto pode permitir verificar possíveis erros existentes na classificação filogenética.
<b>Proposta para Solução</b>	A proposta do projeto consiste em verificar se é possível identificar, com base nas informações dos genes e comparação de genomas, genes mais representativos que vão ao encontro com a classificação filogenética e se, com base nesses genes, é possível identificar a filogenia de maneira mais precisa.
<b>Dados</b>	Os dados utilizados no projeto nos estudos de caso são genomas públicos de bactérias do gênero <i>Xanthomonas</i> , provenientes do banco de genomas do NCBI (National Center for Biotechnology Information). Esse banco contém as sequências de aminoácidos e nucleotídeos desses genomas que serão analisadas e comparadas.
<b>Validação</b>	As ferramentas desenvolvidas e os resultados obtidos serão validados por especialistas do domínio.
<b>Limitações</b>	O projeto estará limitado a estudar e implementar ferramentas para comparar apenas genomas de bactérias. Além disso, a análise das ferramentas depende da comparação dos resultados com a filogenia atual. Por conta disso,

	pode ocorrer um erro na verificação das ferramentas ocasionado por um erro de anotação dessa filogenia.
<b>Resultados esperados</b>	Contribuições científicas: espera-se que a identificação dos genes mais representativos possibilite verificar a anotação de genes que não foram bem feitos, bem como identificar possíveis filogenias de novas espécies.
	Contribuições tecnológicas: será desenvolvido um conjunto de ferramentas computacionais para a comparação entre genomas de bactérias.

<b>MÉTODO DE PESQUISA</b> [para cada item, marque todas as opções adequadas para seu projeto]	
<b>Gênero</b>	<input type="checkbox"/> Pesquisa teórica <input type="checkbox"/> Pesquisa prática <input checked="" type="checkbox"/> Pesquisa empírica <input type="checkbox"/> Pesquisa metodológica
<b>Natureza</b>	<input type="checkbox"/> Pesquisa básica/pura <input checked="" type="checkbox"/> Pesquisa aplicada
<b>Objetivo</b>	<input type="checkbox"/> Pesquisa descritiva <input checked="" type="checkbox"/> Pesquisa exploratória <input type="checkbox"/> Pesquisa explicativa <input type="checkbox"/> [outro – escrever]
<b>Abordagem</b>	<input type="checkbox"/> Pesquisa quantitativa <input type="checkbox"/> Pesquisa qualitativa <input checked="" type="checkbox"/> Pesquisa mista (quali-quant)
<b>Procedimento(s) técnico(s)</b>	<input checked="" type="checkbox"/> Pesquisa experimental <input type="checkbox"/> Pesquisa com <i>survey</i> <input type="checkbox"/> Pesquisa netnográfica <input checked="" type="checkbox"/> Pesquisa bibliográfica <input checked="" type="checkbox"/> Estudo de caso <input type="checkbox"/> Teoria fundamentada em dados ( <i>grounded theory</i> ) <input type="checkbox"/> Pesquisa documental <input type="checkbox"/> Pesquisa participante <input type="checkbox"/> Ciência do projeto ( <i>Design science research</i> ) <input type="checkbox"/> Pesquisa <i>ex-post-facto</i> <input type="checkbox"/> Pesquisa-ação <input type="checkbox"/> [outro – escrever] <input type="checkbox"/> Pesquisa de levantamento <input type="checkbox"/> Pesquisa etnográfica
<b>Fonte(s) de dados</b>	<input checked="" type="checkbox"/> pesquisa de laboratório <input type="checkbox"/> pesquisa de campo <input type="checkbox"/> pesquisa bibliográfica <input checked="" type="checkbox"/> banco de dados de genomas do NCBI
<b>Técnica(s) / Instrumento(s) de coleta de dados</b>	<input type="checkbox"/> medição <input type="checkbox"/> observação (direta / participante) <input type="checkbox"/> questionário <input type="checkbox"/> formulário <input type="checkbox"/> diário de campo / notas de campo <input type="checkbox"/> entrevista <input type="checkbox"/> <i>benchmark</i> <input type="checkbox"/> análise documental (ou de artefatos) <input type="checkbox"/> grupos focais <input checked="" type="checkbox"/> pesquisa em banco de dados de genomas do NCBI
<b>Técnica(s) de análise de dados</b>	Análise quantitativa:     Análise qualitativa: <input type="checkbox"/> Estatística descritiva <input checked="" type="checkbox"/> Análise de conteúdo <input type="checkbox"/> Estatística inferencial <input type="checkbox"/> Análise do discurso <input type="checkbox"/> [outro – escrever] <input type="checkbox"/> [outro – escrever]

<b>CRONOGRAMA</b> [altere conforme necessário]																																							
	2015												2016												2017														
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12			
Estudo da literatura				x	x	x	x	x	x	x	x	x	x	x	x																								
Revisão sistemática							x	x	x																														
Definição da proposta				x	x	x	x																																
Execução da proposta												x	x	x	x	x	x	x																					
Coleta/obtenção de dados												x																											
Validação da proposta																																							
Preparo da qualificação												x	x	x	x																								
Exame de qualificação																																							
Escrita da dissertação																																							
Escrita de artigo																																							
Depósito da dissertação																																							

<sup>i</sup> Esta ficha é uma adaptação da usada no “VIII Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI 2015)” realizado como parte do “XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015)”