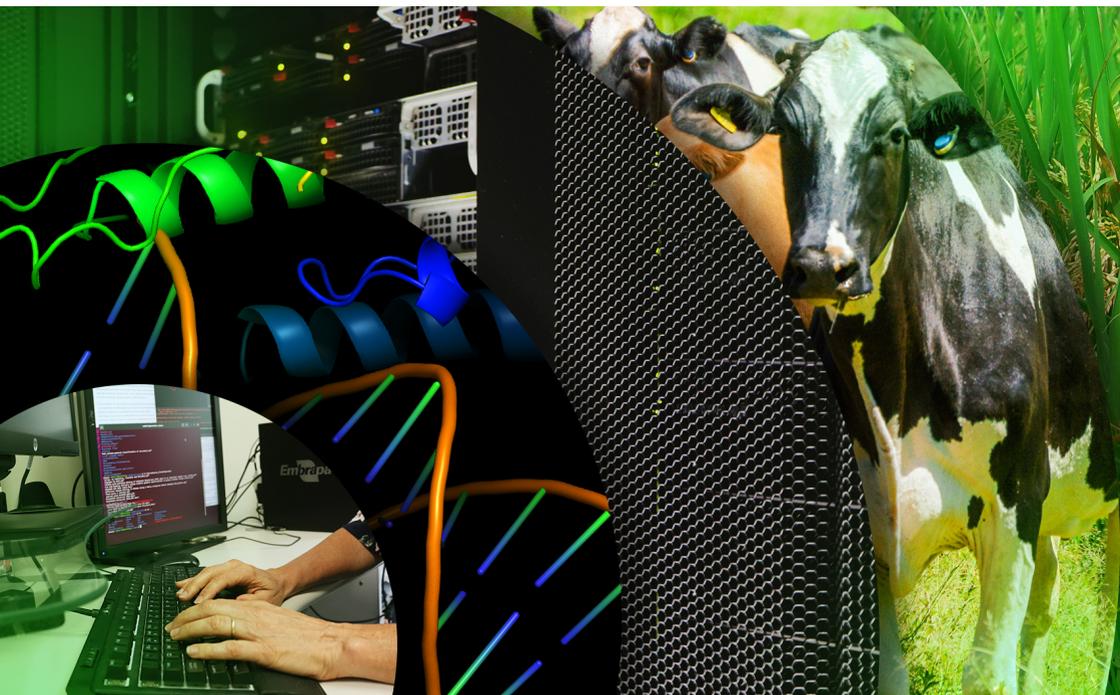


Documentos

Campinas, SP / Outubro, 2024

Laboratório Multiusuário de Bioinformática da Embrapa

Primeiro Decênio



Embrapa

Agricultura Digital

***Empresa Brasileira de Pesquisa Agropecuária
Embrapa Agricultura Digital
Ministério da Agricultura e Pecuária***

e-ISSN 2764-2488

Documentos 191

Outubro, 2024

Laboratório Multiusuário de
Bioinformática da Embrapa

Primeiro Decênio

Paula Regina Kuser Falcão
Michel Eduardo Beleza Yamagishi
Poliana Fernanda Giachetto
Felipe Rodrigues da Silva

***Embrapa Agricultura Digital
Campinas, SP
2024***

Embrapa Agricultura Digital

Av. Dr. André Tosello, 209 - Cidade Universitária
Campinas, SP, Brasil
CEP. 13083-886
www.embrapa.br
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações

Presidente

Júlio Cesar Dalla Mora Esquerdo

Secretária-executiva

Sônia Ternes

Revisão de texto

Graziella Galinari

Normalização bibliográfica

Carla Cristiane Osawa

Projeto gráfico

Leandro Souza Fazio

Diagramação

Lucas Campos Barros e Magda Cruciol

Publicação digital: PDF

Membros

Adauto Luiz Mancini, Alan Massaru Nakai

Carla Cristiane Osawa, Geraldo Cançado,

Graziella Galinari, Joice Machado Bariani,

Juliana Yassitepe, Luiz Manoel Silva Cunha,

Magda Cruciol, Paula Regina Kuser Falcão

Todos os direitos reservados

A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Agricultura Digital

Laboratório Multiusuário de Bioinformática da Embrapa – primeiro decênio / Paula Regina Kuser Falcão ... [et al.]. – Campinas : Embrapa Agricultura Digital, 2024.

PDF (42 p.) : il. color. – (Documentos / Embrapa Agricultura Digital, ISSN 2764-2488 ; 191)

1. Bioinformática. 2. Laboratório multiusuário. 3. Embrapa. I. Falcão, Paula Regina Kuser. II. Embrapa Agricultura Digital. III. Série.

CDD (21. ed.) 570.285

Carla Cristiane Osawa (CRB-8/10421)

© 2024 Embrapa

Autores

Paula Regina Kuser Falcão

Física, doutora em Cristalografia de Proteínas, pesquisadora da Embrapa Agricultura Digital, Campinas, SP.

Michel Eduardo Beleza Yamagishi

Matemático, doutor em Matemática Aplicada e Computacional, pesquisador da Embrapa Agricultura Digital, Campinas, SP.

Poliana Fernanda Giachetto

Zootecnista, doutora em Produção Animal, pesquisadora da Embrapa Agricultura Digital, Campinas, SP.

Felipe Rodrigues da Silva

Biólogo, doutor em Genética e Biologia Molecular, pesquisador da Embrapa Agricultura Digital, Campinas, SP.

Apresentação

A Embrapa é uma empresa de pesquisa fundada em 1973. Fiel a esse propósito, vem, ao longo das décadas, propondo soluções inovadoras para diversos desafios da agricultura nacional. Este trabalho descreve uma dessas soluções inovadoras: o conceito de laboratório multiusuário, especificamente o Laboratório Multiusuário de Bioinformática da Embrapa, conhecido como LMB. Este foi o primeiro laboratório multiusuário implantado na Embrapa e serviu de modelo para os demais que foram criados posteriormente.

A ideia central é oferecer uma infraestrutura de alto custo para projetos de pesquisa, promovendo o compartilhamento racional de recursos limitados. É difícil exagerar a importância dos laboratórios multiusuários, especialmente do LMB. Conforme apresentado neste trabalho, muitos projetos foram viabilizados pelo LMB, diversas publicações científicas produzidas, além de estudantes capacitados. Igualmente relevantes, também foram disponibilizados ao agronegócio patentes e produtos.

Esta publicação é oportuna e necessária, pois celebra o primeiro decênio do LMB e apresenta à sociedade os resultados alcançados. Com orgulho, apresento este texto aos leitores e parabeno todos aqueles que, direta ou indiretamente, contribuíram para o sucesso do LMB.

Stanley Robson de Medeiros Oliveira
Chefe-Geral da Embrapa Agricultura Digital

Sumário

Criação do Laboratório Multiusuário de Bioinformática	12
Infraestrutura computacional e processo de análise de dados	14
Atuação do LMB	18
Considerações finais	22
Referências	24

Introdução

Paula Regina Kuser Falcão

Michel Eduardo Beleza Yamagishi

Poliana Fernanda Giachetto

Felipe Rodrigues da Silva

Em 2023, a Embrapa completou 50 anos. A empresa sempre se destacou por enfrentar os desafios da agropecuária brasileira. Desde sua fundação em 1973, contribuiu para transformar os solos pobres do cerrado brasileiro em terras agricultáveis (Santana et al., 2020). As décadas iniciais também testemunharam o êxito da fixação biológica do nitrogênio, que hoje, só na produção de soja, representa uma economia anual na importação de adubos nitrogenados da ordem de R\$ 38 bilhões (Telles et al., 2023). A chave deste sucesso reside na aplicação do estado da arte científico e métodos avançados para enfrentar os desafios agrícolas do País, como veremos no caso particular dos projetos de pesquisa que demandam o auxílio de ferramentas “ômicas” e análises de bioinformática.

No ano de 1977, Fred Sanger marcou o início da “Era Genômica” ao publicar o genoma do bacteriófago Phi174 (Sanger et al., 1977). No entanto, foi somente a partir de 2001, com a publicação do genoma humano, que essa era se consolidou globalmente (Lander et al., 2001; Venter et al., 2001). Sabe-se que boa parte dos recursos utilizados neste projeto foi empregado na geração dos dados, ou seja, no sequenciamento propriamente dito. Em 2005, no entanto, os sequenciadores Sanger foram superados pelos novos sequenciadores de próxima geração (*Next Generation Sequencing Technologies* (NGS)), reduzindo muito os custos de geração de sequências. Este avanço deslocou o desafio da geração para a análise de dados, demandando aprimoramentos em algoritmos e capacidade computacional para lidar com volumes massivos de sequências curtas.

A Embrapa, sempre atenta às inovações científicas, percebeu rapidamente as mudanças trazidas pela revolução tecnológica e

reconheceu a necessidade de se adaptar a essa disrupção tecnológica. No caminho dessa adaptação, aprovou dois macroprogramas significativos: os projetos Bioinformática e Inovação na Embrapa (BIEM), em 2004, e Pesquisa e Inovação em Bioinformática na Agricultura (PIBA), em 2007, este último agraciado com premiação. Com isso, a Empresa reuniu pesquisadores da área de genômica e de bioinformática para deliberar sobre as demandas em termos de pessoal e de equipamentos, além de realizar um levantamento das necessidades dos diferentes projetos “ômicos” com demanda em bioinformática. Priorizando investimentos em recursos humanos, a Embrapa contratou bioinformatas por meio de concursos públicos, fortalecendo sua equipe interna de pesquisa. Também promoveu o treinamento de seus quadros com a participação em eventos científicos nacionais e internacionais, visitas técnicas e cursos. Numa terceira etapa desta fase de fortalecimento da área, a Embrapa dotou seu sistema de pesquisa com equipamentos de alto desempenho computacional (HPC) para análise de dados genômicos. Como equipamentos de HPC são caros e os projetos individualmente não poderiam arcar com esses custos de investimento, a solução encontrada foi criar uma equipe de bioinformatas e equipar um nodo central com equipamentos de alto desempenho computacional com capacidade de atender aos projetos de pesquisa que usam as tecnologias genômicas e que pudesse ser acessado por todos. Este processo tornou-se a base para o surgimento do Laboratório Multiusuário de Bioinformática da Embrapa (LMB) na Embrapa Agricultura Digital, em 2011.

Criação do Laboratório Multiusuário de Bioinformática

O LMB, estabelecido para enfrentar os desafios da análise de dados genômicos em um ambiente colaborativo, foi criado com uma missão definida. Sua inauguração, em 2011, foi marcada pela presença do então presidente da Embrapa, Maurício Lopes, e foi instituído pela Deliberação nº 55, de 19/09/2011(Embrapa, 2011a) (Figura 1). Por meio da Resolução Normativa nº 16, datada de 22/09/2011 (Embrapa, 2011b), foi aprovado o Plano de Implantação e Operação



Foto: Nadir Rodrigues Pereira

Figura 1. Foto da inauguração do LMB, em outubro de 2011, com a presença do então presidente da Embrapa, Maurício Lopes (à direita na foto); do chefe-geral da Unidade na época, Kléber Sampaio (à esquerda); da atual presidente da Embrapa, Silvia Massruhá, então chefe-adjunta de Pesquisa da Unidade; e do pesquisador Michel Eduardo Beleza Yamagishi (ao centro).

do Laboratório Multiusuário de Bioinformática, em que são apresentados os seus objetivos principais, a criação de um comitê assessor de caráter consultivo e as suas fases de implantação.

O LMB foi criado com o objetivo de viabilizar soluções de bioinformática para projetos de pesquisa, desenvolvimento e inovação na Embrapa em um ambiente colaborativo, atendendo aos seguintes objetivos:

- Contribuir para o avanço na fronteira do conhecimento e incorporar novas tecnologias em bioinformática.
- Viabilizar soluções eficientes para demandas de bioinformática.

- Disponibilizar acesso à infraestrutura computacional de alto desempenho.
- Prover a Embrapa de competências em bioinformática.

O LMB desempenha um papel fundamental como facilitador especializado em projetos que abrangem a geração e análise de sequências genômicas. A bioinformática, dada sua natureza multidisciplinar, atua desde a definição de hipóteses biológicas até a interpretação dos resultados, conferindo ao LMB uma função central na pesquisa agrícola. O processo de análise de dados genômicos implica na minuciosa exploração e interpretação do vasto repositório de informação do genoma de um organismo. Para desvendar os segredos guardados na informação genética de um indivíduo, empregam-se ferramentas avançadas de bioinformática e técnicas computacionais, proporcionando a extração eficaz do conhecimento.

Durante esse processo de extração do conhecimento, são reveladas informações genéticas essenciais. Ao longo de uma década, o LMB tem sido um protagonista nesse cenário, participando em mais de cem projetos de pesquisa e atendendo a mais de 150 usuários de várias instituições. Sua infraestrutura de 130 terabytes de dados biológicos e investimentos contínuos refletem o compromisso em enfrentar os desafios da bioinformática na agricultura.

Infraestrutura computacional e processo de análise de dados

Conforme destacado anteriormente, a bioinformática é inerentemente multidisciplinar, abrangendo uma vasta gama de campos do conhecimento, como engenharia de software, matemática, biologia, agricultura, ciência da computação e tecnologia da informação. Essa área foi concebida para lidar com a enorme quantidade de dados biológicos, antecipando-se ao advento do termo “Big Data”.

No *roadmap* do processo de análises de dados de um projeto de pesquisa em colaboração com o LMB que envolve dados genômicos, descrito na Figura 2, a bioinformática aparece na terceira fase do processo.

Esse *roadmap* revela o caminho desde a formulação de perguntas biológicas até a obtenção de resultados inovadores, envolvendo a aquisição de dados genômicos, pré-processamento, análise e interpretação. Este trajeto, intrinsecamente complexo e interativo, abrange diversos estágios. Inicia-se com a definição de uma hipótese relacionada a um problema ou fenômeno biológico de interesse agrícola. Esse problema pode versar sobre melhoramento genético, doenças, evolução, resistências ou outras áreas da biologia. Uma vez estabelecido o problema, o próximo estágio envolve a aquisição de dados genômicos relevantes, que pode incluir dados de sequenciamento de DNA (genoma completo, RNA-Seq), proteômica, metabolômica, ou outras tecnologias ômicas. A determinação da metodologia de coleta de dados deve preceder a execução, e dados públicos provenientes de fontes como GenBank, Protein Data Bank e ENCODE podem ser inestimáveis fontes complementares de dados.

Assim que uma questão biológica que envolve coleta de dados é feita, a equipe de pesquisadores que trabalha com os organismos conversa com a área de bioinformática para estudar a melhor forma de coleta dos dados. As amostras biológicas são então obtidas para serem enviadas aos laboratórios de geração de dados. Como mencionado acima, podem ser gerados dados de sequenciamento, genotipagem, microarranjos. Esses dados brutos são transferidos para os servidores do Laboratório Multiusuário de Bioinformática, organizados, processados e as análises são realizadas.

Após o recebimento dos dados coletados, a bioinformática passa a ter um papel crítico ao desencadear uma série de etapas de análise e interpretação. Os dados passam por um minucioso processo de pré-processamento que inclui a limpeza dos dados para remoção de ruído e artefatos, além do controle de qualidade. Em casos que envolvem dados multiômicos, ocorre a integração de diferentes conjuntos de dados.

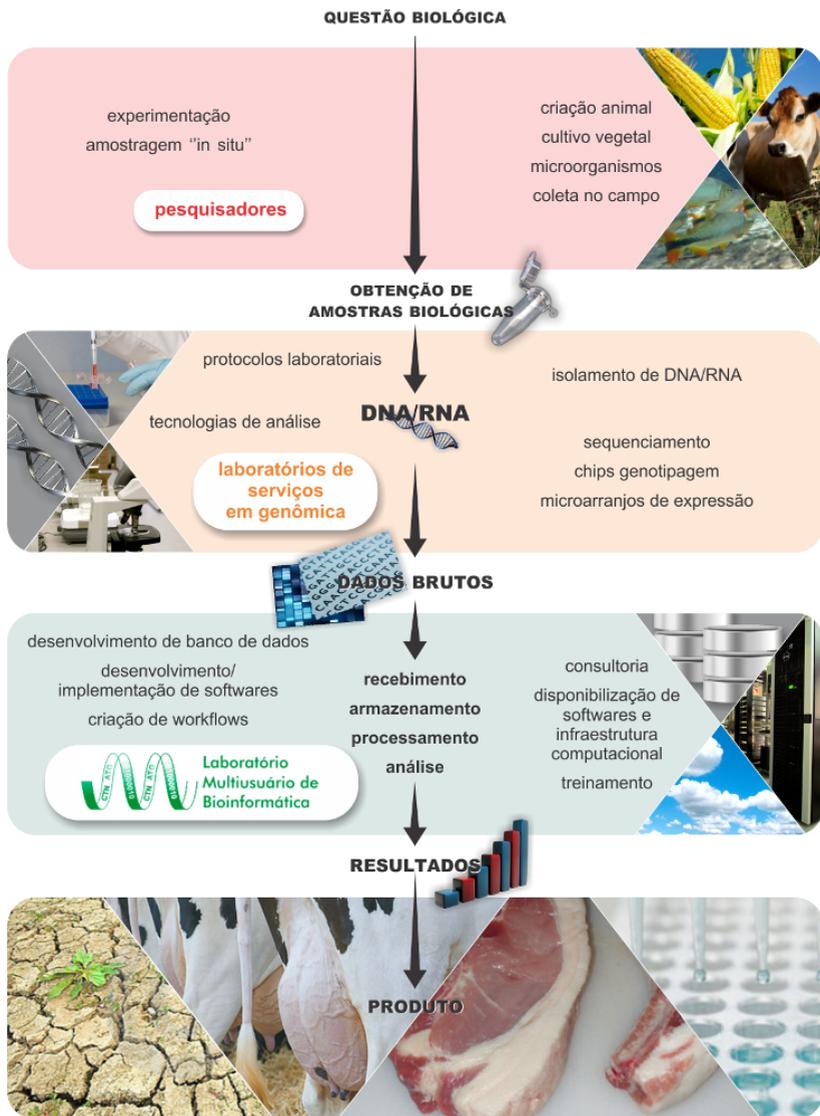


Figura 2. Roadmap dos processos de análise do LMB para projetos geradores de dados biológicos.

A fase analítica compreende o uso de ferramentas de bioinformática e algoritmos especializados, abordando diversas frentes conforme a natureza da questão biológica em estudo. Isso pode incluir busca de variantes, análise da expressão diferencial de genes, investigações filogenéticas e análise de bioinformática estrutural. A aplicação de técnicas estatísticas visa identificar padrões e correlações, fornecendo uma base sólida para a interpretação dos resultados. A visualização desses dados é então realizada para ajudar na exploração dos dados, na comunicação das descobertas e na identificação de *outliers*. A interpretação adequada desses dados pode guiar futuras pesquisas experimentais ou sugerir uma descoberta promissora. Durante todo o caminho percorrido, a colaboração com os cientistas das duas áreas, biologia e bioinformática, é crucial para garantir que a pesquisa é biologicamente significativa e tem um rigor estatístico. O resultado das análises sempre retorna ao pesquisador que gerou a pergunta biológica para que seja aplicada em uma melhoria ou um produto.

O processo que leva desde uma questão biológica até uma inovação significativa é desafiador, mas é imprescindível para descobertas e avanços na agricultura. A equipe do LMB possui uma sólida visão de colaboração para abordar os desafios biológicos de relevância agrícola nas pesquisas da Embrapa por meio da bioinformática, e demonstra este compromisso ao interagir com os diversos programas de pesquisa de melhoramento.

Outras funções do LMB podem envolver o desenvolvimento de bancos de dados para organização dos dados, criação de *workflows* para realização de análises e até mesmo desenvolvimento de software personalizados. O LMB também atua com treinamento de pessoas que desejam fazer suas próprias análises e fornece consultoria sobre as metodologias a serem aplicadas em cada caso.

Na concepção do LMB, foi proposta a manutenção de uma estrutura capaz de atender aos projetos estratégicos demandantes de análises de grande volume de dados por meio de ferramentas especializadas e o fortalecimento de uma equipe de profissionais multidisciplinares, capazes de compreender as questões biológicas envolvidas nos projetos agrícolas e que tivessem formação adequada

para lidar com as análises de dados de grande volume. Para isso, foi criado um plano trienal de investimento cujo objetivo era garantir a atualização de dois aspectos essenciais de qualquer sistema de HPC: o processamento/análise dos dados e o armazenamento seguro destes dados. O processo de análise dos dados proposto para os projetos que utilizam as facilidades do Laboratório Multiusuário de Bioinformática abrange todas as etapas de vida dos dados: recebimento, armazenamento, processamento e análise, tudo isso dentro de uma infraestrutura computacional que suporta todas essas etapas do processo e produz resultados e inovação. Daí a necessidade de manter um sistema HPC atualizado e com garantias de segurança.

Para facilitar o acesso dos usuários à infraestrutura e às análises de bioinformática, implementamos uma política de acesso simples e rápida. Ao mesmo tempo, foi enfatizada a importância da segurança e sigilo dos dados, garantindo um ambiente seguro para as atividades realizadas no laboratório e para os dados armazenados.

Os projetos que são atendidos dentro do LMB envolvem questões que abrangem, entre outros aspectos, a determinação da função de genes, o estudo da evolução de espécies, o aumento da produção, a criação de organismos resistentes e a melhoria da qualidade dos produtos. Para resolver estas questões, os tipos de análises mais trabalhadas pela equipe do LMB são: montagem de genomas e transcriptomas; identificação de genes de interesse; identificação de genes diferencialmente expressos; identificação e caracterização de marcadores moleculares; identificação de SNPs; anotação de genes; análise de qualidade de sequências; análise comparativa de genomas; desenvolvimento de ferramentas.

Atuação do LMB

O modelo organizacional definido para o LMB está descrito na Figura 3. Esse modelo descreve as formas de acesso, entradas e saídas e os relacionamentos estabelecidos.

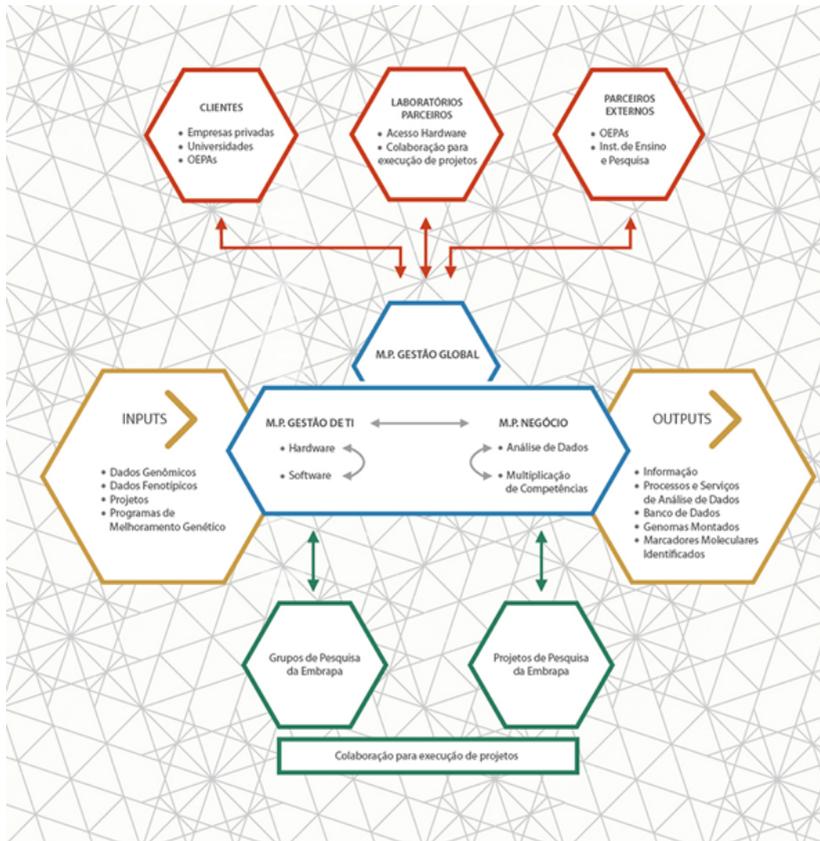


Figura 3. Modelo organizacional do LMB.

O modelo, que foi desenhado na criação do LMB, implicava no atendimento a diversos clientes, desde empresas privadas a parceiros nacionais ou internacionais. E atenderia, principalmente, as pesquisas desenvolvidas na Embrapa colaborando na execução de projetos de interesse da Empresa. O laboratório teria uma gestão dividida entre a parte de infraestrutura e de análise dos dados, e um processo de recebimento de dados e entrega de resultados que se beneficiariam destas duas instâncias.

O acesso ao LMB é feito por meio do site oficial do laboratório, acessível em: <https://www.embrapa.br/agricultura-digital/lmb>. Para utilizar a infraestrutura computacional, é necessário preencher um formulário com algumas informações e um resumo do projeto. Após a análise da solicitação e do plano de atividades descrito, será determinado o período de autorização de acesso. Para garantir um controle adequado de acesso e uso dos servidores, é criado um cadastro com informações do usuário para cada conta de acesso, junto com uma área de trabalho configurada com as permissões apropriadas.

Uma vez que a conta de acesso é criada, o usuário tem acesso remoto ao servidor de submissão de tarefas. Para tarefas que demandam mais tempo de processamento, elas devem ser submetidas por meio de um gestor de filas, o qual define a ordem de execução e gerencia a distribuição de tarefas entre os servidores.

Nossas colaborações em pesquisa englobam a análise de conjuntos de dados volumosos e processamento complexo, empregando diversas ferramentas e protocolos. Frequentemente, isso demanda a repetição de etapas de análise com o ajuste fino de parâmetros. A equipe do LMB já participou ativamente em vários projetos de pesquisa realizando tarefas como montagem de genomas, análise de transcriptomas, análise de expressão diferencial, detecção de SNPs e CNVs, metagenômica, entre outras. Colaboramos para encontrar resultados biologicamente relevantes em projetos distintos, com características diversas e organismos variados, como ovinos, bovinos, peixes, caprinos, arroz, milho, soja, dendê, café, guaraná, eucalipto, banana, trigo, entre outros (Anexo 1).

Ao longo dessa década, a infraestrutura computacional adquirida com recursos do LMB já foi utilizada por mais de 150 usuários de diversas Unidades da Embrapa e outras instituições de pesquisa. Desde a implantação do LMB, foram atendidos mais de cem projetos de pesquisa que demandam processamento de alto desempenho, em colaboração com pelo menos 20 Unidades Descentralizadas de pesquisa da Embrapa (Amazônia Ocidental, Arroz e Feijão, Agrobiologia, Agroenergia, Café, Caprinos e Ovinos, Cerrados, Gado

de Corte, Gado de Leite, Mandioca e Fruticultura, Meio Ambiente, Milho e Sorgo, Pecuária Sudeste, Pecuária Sul, Pesca e Aquicultura, Recursos Genéticos, Semiárido, Soja, Suínos e Aves, Trigo) e também com instituições externas, como o Instituto de Biologia da Universidade Estadual de Campinas (IB-Unicamp), Escola Superior de Agricultura “Luiz de Queiroz” (Esalq), Universidade Estadual Paulista (Unesp), Universidade Federal de Minas Gerais (UFMG), Universidade Federal de Pelotas (UFPel), Universidade Federal do Rio Grande do Sul (UFRGS), Universidade Federal de Mato Grosso do Sul (UFMS), Universidade de São Paulo (USP), Universidade do Estado de Mato Grosso (Unemat), Centro de Aquicultura da Unesp (Caunesp), Universidade Federal de São Carlos (UFSCar), Universidade do Estado de Santa Catarina (Udesc), Comissão de Biossegurança do Instituto de Biologia (CBio-IB) da Universidade Federal Fluminense (UFF), Instituto Agrônomico de Pernambuco (IPA), Universidade de Wageningen (Holanda), Biotechnology and Biological Sciences Research Council’s (UK), Universidade de Cornell (USA). As publicações, patentes e produtos resultantes destes trabalhos pode ser encontrada no Anexo 2.

No que diz respeito à formação de recursos humanos em bioinformática, a equipe do laboratório contribuiu diretamente por meio de:

- Treinamentos em ferramentas de bioinformática.
- Orientação/coorientação de dissertações de mestrado e teses de doutorado para alunos de várias universidades.
- Recebimento de visitantes para treinamentos de curta duração (Anexo 3).

Estas atividades de formação já beneficiaram ao menos 400 usuários, resultando na capacitação de profissionais que atualmente trabalham no Brasil e no exterior. Além disso, em parceria com o programa de pós-graduação do Instituto de Biologia da Universidade Estadual de Campinas, foram ministrados cursos específicos em bioinformática para colegas da Embrapa e parceiros externos nos

últimos anos. O laboratório também estabeleceu parceria com a instituição EMBL-EBI (European Bioinformatics Institute) no projeto CABANA (Capacity Building for Bioinformatics in Latin America), disponível em: <http://cabana.online>.

Considerações finais

O sucesso alcançado pelo LMB ao longo desta década é atribuído, em grande parte, à sua equipe altamente qualificada, ao investimento em máquinas de computação de alto desempenho e à criação de uma extensa rede de colaboradores nacionais e internacionais. Desde sua criação, a equipe do Laboratório Multiusuário de Bioinformática da Embrapa liderou e contribuiu em projetos de grande relevância, tendo seu reconhecimento reafirmado por meio de diversas premiações, como a mencionada no Projeto PIBA. Além desta, teve ainda a premiação nacional do sistema Embrapa nas categorias “Parcerias” e “Qualidade Técnica” com o projeto Rede Genômica Animal; a premiação nacional do sistema Embrapa na categoria “Equipes - Ação Gerencial Categoria Parcerias”, em 2012; e premiação nacional de equipes na categoria “Criatividade”, com o projeto Identificação de Genes de Interesse para a Suinocultura por meio de Genotipagem SNPs em Grande Escala. Também teve destaque no aniversário de 47 anos da Embrapa com a tecnologia TambaPlus e, nas comemorações de 48 anos da Empresa, com a Plataforma AquaPlus. Ainda, em 2020, um membro da equipe, o pesquisador Michel Yamagishi, recebeu o prêmio inventores Unicamp - Categoria Patentes Concedidas, da Agência de Inovação da Universidade Estadual de Campinas (Inova Unicamp).

Ao longo dos anos, a bioinformática passou por uma evolução notável, desencadeando transformações significativas que a impulsionaram para uma nova e empolgante era. O crescimento exponencial no volume e complexidade dos dados biológicos, impulsionado pelos avanços em tecnologias de sequenciamento de alto rendimento, proteômica, metabolômica e outros campos “ômicos”, demanda ferramentas e abordagens computacionais

sofisticadas para a análise eficaz e interpretação desses conjuntos de dados. Essa revolução resultou na integração da pesquisa biológica tradicional com métodos computacionais, gerando uma colaboração interdisciplinar que se tornou a marca registrada da bioinformática moderna. Os limites entre a biologia e a ciência da computação estão se tornando cada vez mais difusos, os bioinformatas de hoje lideram essa convergência, preenchendo cada vez mais a lacuna entre duas disciplinas aparentemente distintas.

Após completar uma década de existência, o Laboratório Multiusuário de Bioinformática encontra-se num ambiente onde a bioinformática difere da bioinformática de dez anos atrás em vários aspectos importantes, principalmente devido aos avanços tecnológicos e às mudanças na disponibilidade de dados. Nos últimos dez anos, houve avanços significativos nas tecnologias de sequenciamento de DNA e RNA. O sequenciamento de próxima geração (NGS) tornou-se mais acessível. Com isso, houve um crescimento no volume e na complexidade dos dados biológicos disponíveis, exigindo o desenvolvimento de ferramentas computacionais mais sofisticadas e tornando a análise de dados um elemento essencial em todo o processo da pesquisa biológica.

O aumento na integração de diferentes tipos de dados biológicos, como genômica, transcriptômica, proteômica e metabolômica levou a uma abordagem multiômica que permite uma compreensão mais abrangente dos processos biológicos, mas também aumenta os desafios computacionais. Abordagens com técnicas de *machine learning* e inteligência artificial têm sido usadas cada vez mais. A bioinformática hoje é caracterizada por um ambiente de pesquisa dinâmico e laboratórios como o LMB precisam lidar com desafios computacionais complexos associados à análise integrada de grandes conjuntos de dados, exigindo um aumento considerável da capacidade computacional e da infraestrutura de armazenamento de dados. Antevendo esta necessidade, a infraestrutura disponibilizada pelo LMB deverá em breve ser incrementada com a aquisição de novas máquinas de processamento e armazenamento com

financiamento público (projeto da Finep, chamada pública MCTI/FINEP/CT-Infra 04/2018).

Com uma infraestrutura adequada, uma equipe coesa e interdisciplinar, investimento em treinamento e na organização padronizada dos dados gerados a bioinformática tem o potencial de promover avanços significativos na agricultura.

Referências

EMBRAPA. Deliberação nº 55, de 19 de setembro de 2011. [Institui o Laboratório Multiusuário de Bioinformática, integrado às competências já existentes na Embrapa]. **Boletim de Comunicações Administrativas**, ano 37, n. 50, p. 7-8, 24 out. 2011a.

EMBRAPA. Resolução Normativa nº 16, de 22 de setembro de 2011. [Aprova o anexo “Plano de Implantação e Operação do Laboratório Multiusuário de Bioinformática da Embrapa”, instalado na Embrapa Informática Agropecuária]. **Boletim de Comunicações Administrativas**, ano 37, n. 50, p. 9, 24 out. 2011b.

LANDER, E. S.; LINTON, L. M.; BIRREN, B.; NUSBAUM, C.; ZODY, M. C.; BALDWIN, J. DEVON, K.; DEWAR, K.; DOYLE, M.; FITZHUGH, W.; FUNKE, R.; GAGE, D.; HARRIS, K.; HEAFORD, A.; HOWLAND, J.; KANN, L.; LEHOCZKY, J.; LEVINE, R.; MCEWAN, P.; MCKERNAN, K.; MELDRIM, J.; MESIROV, J. P.; MIRANDA, C.; MORRIS, W.; NAYLOR, J.; RAYMOND, C.; ROSETTI, M.; SANTOS, R.; SHERIDAN, A.; SOUGNEZ, C.; STANGE-THOMANN, N.; STOJANOVIC, N.; SUBRAMANIAN, A.; WYMAN, D.; ROGERS, J.; SULSTON, J.; AINSCOUGH, R.; BECK, S.; BENTLEY, D.; BURTON, J.; CLEE, C.; CARTER, N.; COULSON, A.; DEADMAN, R.; DELOUKAS, P.; DUNHAM, A.; DUNHAM, I.; DURBIN, R.; FRENCH, L.; GRAFHAM, D.; GREGORY, S.; HUBBARD, T.; HUMPHRAY, S.; HUNT, A.; JONES, M.; LLOYD, C.; MCMURRAY, A.; MATTHEWS, L.; MERCER, S.; MILNE, S.; MULLIKIN, J. C.; MUNGALL, A.; PLUMB, R.; ROSS, M.; SHOWNKEEN, R.; SIMS, S.; WATERSTON, R. H.; WILSON, R. K.; HILLIER, L. W.; MCPHERSON, J. D.; MARRA, M. A.; MARDIS, E. R.; FULTON, L. A.; CHINWALLA, A. T.; PEPIN, K. H.; GISH, W. R.; CHISSOE, S. L.; WENDL, M. C.; DELEHAUNTY, K. D.; MINER, T. L.; DELEHAUNTY, A.; KRAMER, J. B.; COOK, L. L.; FULTON, R. S.; JOHNSON, D. L.; MINX,

P. J.; CLIFTON, S. W.; HAWKINS, T.; BRANSCOMB, E.; PREDKI, P.; RICHARDSON, P.; WENNING, S.; SLEZAK, T.; DOGGETT, N.; CHENG, J. F.; OLSEN, A.; LUCAS, S.; ELKIN, C.; UBERBACHER, E.; FRAZIER, M.; GIBBS, R. A.; MUZNY, D. M.; SCHERER, S. E.; BOUCK, J. B.; SODERGREN, E. J.; WORLEY, K. C.; RIVES, C. M.; GORRELL, J. H.; METZKER, M. L.; NAYLOR, S. L.; KUCHERLAPATI, R. S.; NELSON, D. L.; WEINSTOCK, G. M.; SAKAKI, Y.; FUJIYAMA, A.; HATTORI, M.; YADA, T.; TOYODA, A.; ITOH, T.; KAWAGOE, C.; WATANABE, H.; TOTOKI, Y.; TAYLOR, T.; WEISSENBACH, J.; HEILIG, R.; SAURIN, W.; ARTIGUENAVE, F.; BROTTIER, P.; BRULS, T.; PELLETIER, E.; ROBERT, C.; WINCKER, P.; SMITH, D. R.; DOUCETTE-STAM, L.; RUBENFIELD, M.; WEINSTOCK, K.; LEE, H. M.; DUBOIS, J.; ROSENTHAL, A.; PLATZER, M.; NYAKATURA, G.; TAUDIEN, S.; RUMP, A.; YANG, H. M.; YU, J.; WANG, J.; HUANG, G. Y.; GU, J.; HOOD, L.; ROWEN, L.; MADAN, A.; QIN, S.; DAVIS, R. W.; FEDERSPIEL, N. A.; ABOLA, A. P.; PROCTOR, M. J.; MYERS, R. M.; SCHMUTZ, J.; DICKSON, M.; GRIMWOOD, J.; COX, D. R.; OLSON, M. V.; KAUL, R.; RAYMOND, C.; SHIMIZU, N.; KAWASAKI, K.; MINOSHIMA, S.; EVANS, G. A.; ATHANASIOU, M.; SCHULTZ, R.; ROE, B. A.; CHEN, F.; PAN, H. Q.; RAMSER, J.; LEHRACH, H.; REINHARDT, R.; MCCOMBIE, W. R.; BASTIDE, M. de la; DEDHIA, N.; BLÖCKER, H.; HORNISCHER, K.; NORDSIEK, G.; AGARWALA, R.; ARAVIND, L.; BAILEY, J. A.; BATEMAN, A.; BATZOGLOU, S.; BIRNEY, E.; BORK, P.; BROWN, D. G.; BURGE, C. B.; CERUTTI, L.; CHEN, H. C.; CHURCH, D.; CLAMP, M.; COPLEY, R. R.; DOERKS, T.; EDDY, S. R.; EICHLER, E. E.; FUREY, T. S.; GALAGAN, J.; GILBERT, J. G. R.; HARMON, C.; HAYASHIZAKI, Y.; HAUSSLER, D.; HERMJAKOB, H.; HOKAMP, K.; JANG, W.; JOHNSON, L. S.; JONES, T. A.; KASIF, S.; KASPRYZK, A.; KENNEDY, S.; KENT, W. J.; KITTS, P.; KOONIN, E. V.; KORF, I.; KULP, D.; LANCET, D.; LOWE, T. M.; MCLYSAGHT, A.; MIKKELSEN, T.; MORAN, J. V.; MULDER, N.; POLLARA, V. J.; PONTING, C. P.; SCHULER, G.; SCHULTZ, J.; SLATER, G.; SMIT, A. F. A.; STUPKA, E.; SZUSTAKOWKI, J.; THIERRY-MIEG, D.; THIERRY-MIEG, J.; WAGNER, L.; WALLIS, J.; WHEELER, R.; WILLIAMS, A.; WOLF, Y. I.; WOLFE, K. H.; YANG, S. P.; YEH, R. F.; COLLINS, F.; GUYER, M. S.; PETERSON, J.; FELSENFELD, A.; WETTERSTRAND, K. A.; PATRINOS, A.; MORGAN, M.

J. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860–921, Feb. 2001. DOI: <https://doi.org/10.1038/35057062>.

SANGER, F.; AIR, G. M.; BARRELL, B. G.; BROWN, N. L.; COULSON, A. R.; FIDDES, J. C.; HUTCHISON III, C. A.; SLOCOMBE, P. M.; SMITH, M. Nucleotide sequence of bacteriophage Φ X174 DNA. **Nature**, v. 265, n. 5596 p. 687-95, Feb. 1977. DOI: 10.1038/265687a0.

SANTANA, C. A. M.; CAMPOS, S. K.; MARRA, R.; ARAGÃO, A. A. Cerrado: pilar da agricultura brasileira. In: BOLFE, E. L.; SANO, E. E.; CAMPOS, S. K. (ed.). **Dinâmica agrícola no cerrado: análises e projeções**. Brasília, DF: Embrapa, 2020. v. 1, cap. 2, p. 39-58. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1121866>. Acesso em: 9 ago. 2024.

TELLES, T. S.; NOGUEIRA, M. A.; HUNGRIA, M. Economic value of biological nitrogen fixation in soybean crops in Brazil. **Environmental Technology & Innovation**, v. 31, 103158, Aug. 2023. DOI: <https://doi.org/10.1016/j.eti.2023.103158>.

VENTER, J. C.; ADAMS, M. D.; MYERS, E. W.; LI, P. W.; MURAL, R. J.; SUTTON, G. G.; SMITH, H. O.; YANDELL, M.; EVANS, C. A.; HOLT, R. A.; GOCAYNE, J. D.; AMANATIDES, P.; BALLEW, R. M.; HUSON, D. H.; WORTMAN, J. R.; ZHANG, Q.; KODIRA, C. D.; ZHENG, X. H.; CHEN, L.; SKUPSKI, M.; SUBRAMANIAN, G.; THOMAS, P. D.; ZHANG, J.; GABOR MIKLOS, G. L.; NELSON, C.; BRODER, S.; CLARK, A. G.; NADEAU, J.; MCKUSICK, V. A.; ZINDER, N.; LEVINE, A. J.; ROBERTS, R. J.; SIMON, M.; SLAYMAN, C.; HUNKAPILLER, M.; BOLANOS, R.; DELCHER, A.; DEW, I.; FASULO, D.; FLANIGAN, M.; FLOREA, L.; HALPERN, A.; HANNENHALLI, S.; KRAVITZ, S.; LEVY, S.; MOBARRY, C.; REINERT, K.; REMINGTON, K.; ABU-THREIDEH, J.; BEASLEY, E.; BIDDICK, K.; BONAZZI, V.; BRANDON, R.; CARGILL, M.; CHANDRAMOULISWARAN, I.; CHARLAB, R.; CHATURVEDI, K.; DENG, Z.; DI FRANCESCO, V.; DUNN, P.; EILBECK, K.; EVANGELISTA, C.; GABRIELIAN, A. E.; GAN, W.; GE, W.; GONG, F.; GU, Z.; GUAN, P.; HEIMAN, T. J.; HIGGINS, M. E.; JI, R. R.; KE, Z.; KETCHUM, K. A.; LAI, Z.; LEI, Y.; LI, Z.; LI, J.; LIANG, Y.; LIN, X.; LU, F.; MERKULOV, G. V.; MILSHINA, N.; MOORE, H. M.; NAIK, A. K.; NARAYAN, V. A.; NEELAM, B.; NUSSKERN, D.; RUSCH, D. B.; SALZBERG, S.; SHAO, W.; SHUE, B.; SUN, J.; WANG, Z.; WANG, A.; WANG, X.; WANG, J.; WEI, M.; WIDES, R.; XIAO, C.; YAN, C.; YAO, A.; YE, J.; ZHAN, M.; ZHANG, W.; ZHANG,

H.; ZHAO, Q.; ZHENG, L.; ZHONG, F.; ZHONG, W.; ZHU, S.; ZHAO, S.; GILBERT, D.; BAUMHUETER, S.; SPIER, G.; CARTER, C.; CRAVCHIK, A.; WOODAGE, T.; ALI, F.; AN, H.; AWE, A.; BALDWIN, D.; BADEN, H.; BARNSTEAD, M.; BARROW, I.; BEESON, K.; BUSAM, D.; CARVER, A.; CENTER, A.; CHENG, M. L.; CURRY, L.; DANAHER, S.; DAVENPORT, L.; DESILETS, R.; DIETZ, S.; DODSON, K.; DOUP, L.; FERRIERA, S.; GARG, N.; GLUECKSMANN, A.; HART, B.; HAYNES, J.; HAYNES, C.; HEINER, C.; HLADUN, S.; HOSTIN, D.; HOUCK, J.; HOWLAND, T.; IBEGWAM, C.; JOHNSON, J.; KALUSH, F.; KLINE, L.; KODURU, S.; LOVE, A.; MANN, F.; MAY, D.; MCCAWLEY, S.; MCINTOSH, T.; MCMULLEN, I.; MOY, M.; MOY, L.; MURPHY, B.; NELSON, K.; PFANNKOCH, C.; PRATTS, E.; PURI, V.; QURESHI, H.; REARDON, M.; RODRIGUEZ, R.; ROGERS, Y. H.; ROMBLAD, D.; RUHFEL, B.; SCOTT, R.; SITTER, C.; SMALLWOOD, M.; STEWART, E.; STRONG, R.; SUH, E.; THOMAS, R.; TINT, N. N.; TSE, S.; VECH, C.; WANG, G.; WETTER, J.; WILLIAMS, S.; WILLIAMS, M.; WINDSOR, S.; WINN-DEEN, E.; WOLFE, K.; ZAVERI, J.; ZAVERI, K.; ABRIL, J. F.; GUIGÓ, R.; CAMPBELL, M. J.; SJOLANDER, K. V.; KARLAK, B.; KEJARIWAL, A.; MI, H.; LAZAREVA, B.; HATTON, T.; NARECHANIA, A.; DIEMER, K.; MURUGANUJAN, A.; GUO, N.; SATO, S.; BAFNA, V.; ISTRAIL, S.; LIPPERT, R.; SCHWARTZ, R.; WALENZ, B.; YOOSEPH, S.; ALLEN, D.; BASU, A.; BAXENDALE, J.; BLICK, L.; CAMINHA, M.; CARNES-STINE, J.; CAULK, P.; CHIANG, Y. H.; COYNE, M.; DAHLKE, C.; DESLATTES MAYS, A.; DOMBROSKI, M.; DONNELLY, M.; ELY, D.; ESPARHAM, S.; FOSLER, C.; GIRE, H.; GLANOWSKI, S.; GLASSER, K.; GLODEK, A.; GOROKHOV, M.; GRAHAM, K.; GROPMAN, B.; HARRIS, M.; HEIL, J.; HENDERSON, S.; HOOVER, J.; JENNINGS, D.; JORDAN, C.; JORDAN, J.; KASHA, J.; KAGAN, L.; KRAFT, C.; LEVITSKY, A.; LEWIS, M.; LIU, X.; LOPEZ, J.; MA, D.; MAJOROS, W.; MCDANIEL, J.; MURPHY, S.; NEWMAN, M.; NGUYEN, T.; NGUYEN, N.; NODELL, M.; PAN, S.; PECK, J.; PETERSON, M.; ROWE, W.; SANDERS, R.; SCOTT, J.; SIMPSON, M.; SMITH, T.; SPRAGUE, A.; STOCKWELL, T.; TURNER, R.; VENTER, E.; WANG, M.; WEN, M.; WU, D.; WU, M.; XIA, A.; ZANDIEH, A.; ZHU, X. The sequence of the human genome. **Science**, v. 291, n. 5507, p. 1304-1351, Feb. 2001. DOI:<https://doi.org/10.1126/science.1058040>.

Anexo 1 — Projetos desenvolvidos no LMB nos seus primeiros dez anos de existência¹

- Abordagens moleculares integrativas aplicáveis ao desenvolvimento de ativos tecnológicos que reduzam a emissão de gases de efeito estufa por bovinos de corte.
- Análise do transcriptoma de *Paspalum spp.* sob estresse hídrico.
- Aquagen: Desenvolvimento e validação de ferramentas genômicas aplicadas à aquíicultura.
- Inibidores da poligalacturonase de fungos fitopatogênicos: uma forma de controle da monília e da vassoura-de-bruxa e em cacau e cupuaçu.
- BioSSecco: biologia de sistemas de adaptação à seca, calor e alta concentração de CO₂ em *Vellozia spp.*
- Carrapato bovino: Identificação de antígenos para o desenvolvimento de vacina através de vacinologia reversa.
- Comparação dos genomas bovino e taurino.
- Construção de redes gênicas a partir de dados de microarranjos.
- Desenvolvimento e validação de marcadores SNPs para identificação de cultivares e rastreabilidade de cafés especiais.
- Exploração do microbioma da rizosfera para uma produção sustentável do trigo.
- Expressão diferencial em seringueiras.
- Hologenoma de Nelore: implicações na qualidade de carne e em eficiência alimentar.
- Identificação de CNVs e associação com características produtivas em Canchim.
- Identificação de genes em cana-de-açúcar, relacionados a tolerância da planta a condições de seca prolongada.
- Identificação de genes em caprinos, em busca do mecanismo de tolerância dos animais ao parasita gastrointestinal *Haemonchus contortus*.
- Identificação de genes em diferentes raças e cruzamentos de gado de corte, envolvidos com qualidade da carne e tolerância ao carrapato bovino (*Rhipicephalus microplus*).
- Identificação de genes em eucalipto para o melhoramento genético da madeira para a indústria de papel e celulose.
- Identificação de genes em trigo, relacionados a resistência ao fungo causador da ferrugem (*Puccinia triticina*).
- Identificação de marcadores moleculares relacionados a características reprodutivas e de desempenho em suínos.
- Interação genótipo x ambiente na qualidade dos grãos e bebida em café.
- JM-CNV: desenvolvimento de software para a identificação de CNVRs (regiões de Copy Number Variation).
- KOMODO: desenvolvimento de software para detecção de genes homólogos, significativamente sub ou super representados entre táxons.
- Machado: framework para armazenar, buscar e visualizar informação biológica.

¹ Os projetos listados tiveram liderança de membros da equipe do LMB e também parceria com outras Unidades da Embrapa e instituições.

- Maxiplat: Plataforma computacional para integrar e armazenar os dados e informações provenientes dos projetos do arranjo MaxiBife relativos aos fenótipos, pedigree e genótipos dos diversos rebanhos, bem como dados econômicos e ambientais inerentes aos sistemas produtivos.
- Microbioma da rizosfera de bananeira na ausência e presença do fusarium da banana (*Fusarium oxysporum* f.sp. *cubense*).
- Montagem do genoma da Cachara (*Pseudoplatystoma fasciatum*) testando nova tecnologia de montagem.
- Montagem do genoma do caiaué (*Elaeis oleifera*), espécie de oleaginosa da família do dendê, considerada uma das promessas para o aumento de produtividade a campo relacionada a produção de óleo.
- Montagem do genoma do camarão para estudo populacional com a utilização de técnica de redução de complexidade para digerir partes repetitivas. (<https://www.embrapa.br/busca-de-noticias/-/noticia/60909107/ferramenta-genetica-faz-analise-de-linhagens-de-camarao-cinza>)
- Montagem de genoma de duas cepas de lentivírus de cabra.
- Montagem do genoma do guaraná.
- Montagem do genoma Nelore (*Bos taurus indicus*), como subsídio para a realização de pesquisas que buscam melhorias em termos de produtividade e qualidade de carne na raça bovina de maior representatividade na pecuária de corte nacional.
- Montagem do genoma do Tambaqui (*Colossoma macropomum*), espécie aquícola nativa mais importante do Brasil, constituindo uma fonte de informação para a identificação de marcadores moleculares, melhoramento genético, conservação e estudos biológicos da espécie.
- PlantAnnot, Plant Co-expression Annotation Resource. Sistema desenvolvido para encontrar proteínas que não têm nenhuma anotação ou função atribuída (PUFs). (<https://www.machado.cnptia.embrapa.br/plantannot>)
- POTION: desenvolvimento de software para detecção de seleção positiva e genes envolvidos em processos adaptativos.
- Programa de melhoramento do trigo: marcadores moleculares para diversas características de interesse em trigo.
- RNAseq de bactérias *Aeromonas hydrophila* e *Streptococcus agalactiae*, resistentes ou não ao antibiótico florfenicol.
- Tecnologias para computação distribuída, armazenamento de grandes volumes de dados e workflow científico, em suporte à pesquisa agropecuária.
- Tolerância à seca prolongada em duas cultivares de cana-de-açúcar.
- Vírus do Mosaico: Estudo de variabilidade genética em diferentes cultivares de trigo.
-

Anexo 2 – Publicações de trabalhos realizados utilizando recursos do LMB

2012

BERG, A. S.; DITA, M.; NAN, T.; SHEA, T.; ZHOU, S.; JONKERS, W.; ZHENG, Q.; YOUNG, S.; YAMAGISHI, M. E. B.; GIACHETTO, P. F.; HERAI, R.; SOUZA, M.; WAALWIJK, C.; KEMA, G. H. J.; KLITSLER, H.; MA, L. Genome sequencing of *Fusarium oxysporum* f. sp. cubense tropical race 4 strain II5. **Phytopathology**, v. 102, n. 7, 2012. Supplement 4. Abstracts of presentations of the APS Annual meeting, Providence, RI, 2012.

BOARETO, M.; YAMAGISHI, M. E. B.; CATICHA, N.; LEITE, V. B. P. Relationship between global structural parameters and Enzyme Commission hierarchy: implications for function prediction. **Computational Biology and Chemistry**, v. 40, p. 15-19, 2012. DOI: <https://doi.org/10.1016/j.compbiolchem.2012.06.003>.

FARIA, L. C. B.; ROCHA, A. S. L.; KLEINSCHMIDT, J. H.; SILVA-FILHO, M. C.; BIM, E.; HERAI, R. H.; YAMAGISHI, M. E. B.; PALAZZO JÚNIOR, R. Is a genome a codeword of an error-correcting code? **PLoS One**, v. 7, n. 5, e105396, May 2012. DOI: <https://doi.org/10.1371/journal.pone.0036644>.

2013

DITA, M. A.; HERAI, R. H.; WAALWIJK, C.; YAMAGISHI, M. E. B.; GIACHETTO, P. F.; FERREIRA, G.; SOUZA, M.; KEMA, G. H. J. Comparative transcriptome analyses and genome assembly of *Fusarium oxysporum* f. sp. cubense. **Acta Horticulturae**, v. 986, p. 165-168, 2013.

YAMAGISHI, M. E. B.; HERAI, R. H. **Expanding the Grammar of Biology**. 2013. Disponível em: <https://arxiv.org/abs/1312.3350>. Acesso em: 9 ago. 2024.

2014

CARDOSO, D. C.; MARTINATI, J. C.; GIACHETTO, P. F.; VIDAL, R. O.; CARAZZOLLE, M. F.; PADILHA, L.; GUERREIRO-FILHO, O.; MALUF, M. Large-scale analysis of differential gene expression in coffee genotypes resistant and susceptible to leaf miner-toward the identification of candidate genes for marker assisted-selection. **BMC Genomics**, v. 15, 66, 2014. DOI: <https://doi.org/10.1186/1471-2164-15-66>.

GRATIVOL, C.; REGULSKI, M.; BERTALAN, M.; MCCOMBIE, W. R.; DA SILVA, F. R. da; ZERLOTINI NETO, A.; VICENTINI, R.; FARINELLI, L.; HEMERLY, A. S.; MARTIENSSEN, R. A.; FERREIRA, P. C. G. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. **The Plant Journal**, v. 79, n. 1, p. 162-172, 2014. DOI: <https://doi.org/10.1111/tbj.12539>.

2015

BAMBINI, M. D.; GIACHETTO, P. F.; FALCÃO, P. R. K.; OLIVEIRA, F. S. de. Emerging biotechnologies: bioinformatics services applied to agriculture. In: CONGRESSO LATINO-IBEROAMERICANO DE GESTÃO DA TECNOLOGIA, 16., 2015, Porto Alegre. **Inovação para além da tecnologia**: anais. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2015. Não paginado. Altec 2015. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1027409>. Acesso em: 9 ago. 2024.

CRUZ, A. S. da. **Estudo de associação ampla do genoma bovino para lactação ajustada em 305 dias em Girolando**. 2015. 113 p. Tese (Doutorado Ciências Biológicas) — Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia.

DIAS, Z.; DIAS, U. Sorting by prefix reversals and prefix transpositions. **Discrete Applied Mathematics**. v. 181, p. 78-89, Jan. 2015. DOI: <https://doi.org/10.1016/j.dam.2014.09.004>.

FONSECA, I.; CARDOSO, F. F.; HIGA, R. H.; GIACHETTO, P. F.; BRANDÃO, H. de M.; BRITO, M. A. V. P.; FERREIRA, M. B. D.; GUIMARÃES, S. E. F.; MARTINS, M. F. Gene expression profile in zebu dairy cows (*Bos taurus indicus*) with mastitis caused by *Streptococcus agalactiae*. **Livestock Science**, v. 180, p. 47-57, Oct. 2015. DOI: <https://doi.org/10.1016/j.livsci.2015.07.012>.

HEBERLE, H.; MEIRELLES, G. V.; SILVA, F. R. da; TELLES, G. P.; MINGHIM, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. **BMC Bioinformatics**, v. 16, 169, 2015. DOI: <https://doi.org/10.1186/s12859-015-0611-3>.

HONGO, J. A.; CASTRO, G. M. de; CINTRA, L. C.; ZERLOTINI, A.; LOBO, F. P. POTION: an end-to-end pipeline for positive Darwinian selection detection in genome-scale data through phylogenetic comparison of protein-coding genes. **BMC Genomics**, v. 16, n. 1, 567, 2015. DOI: <https://doi.org/10.1186/s12864-015-1765-0>.

SILVA, J. M. da; GIACHETTO, P. F.; SILVA, L. O. C. da; CINTRA, L. C.; PAIVA, S. R.; CAETANO, A. R.; YAMAGISHI, M. E. B. Genomic variants revealed by invariably missing genotypes in Nelore cattle. **PLoS One**, v. 10, n. 8, e0136035, 2015. DOI: <https://doi.org/10.1371/journal.pone.0136035>.

SILVA, V. H. da; REGITANO, L. C. de A; GEISTLINGER, L.; PÉRTILLE, F.; GIACHETTO, P. F.; BRASSALOTI, R. A.; MOROSINI, N. S.; ZIMMER, R.; COUTINHO, L. L. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. **PLoS One**, v. 11, n. 6, e0157711, 2016. DOI: <https://doi.org/10.1371/journal.pone.0157711>.

VIEIRA, F. D.; OLIVEIRA, S. R. de M.; PAIVA, S. R. Metodologia baseada em técnicas de mineração de dados para suporte à certificação de raças de ovinos. **Engenharia Agrícola**, v. 35, n. 6, p. 1172-1186, nov./dez. 2015. DOI: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v35n6p1172-1186/2015>.

2016

CASTRO, G. M. de; LIMA, M. F.; FONSECA, M. E. N.; BOITEUX, L. S.; LOBO, F. P.; SILVA, F. R.; RECH, E. L. First report of papaya ringspot virus-type W infecting *Feuillea* species (Cucurbitaceae) in South America. **Plant Disease**, v. 100, n. 12, p. 2240, Dec. 2016. DOI: <http://dx.doi.org/10.1094/PDIS-05-16-0662-PDN>.

CSORDAS, B. G.; GARCIA, M. V.; CUNHA, R. C.; GIACHETTO, P. F.; BLECHA, I. M. Z.; ANDREOTTI, R. New insights from molecular characterization of the tick *Rhipicephalus* (*Boophilus*) *microplus* in Brazil. **Revista Brasileira de Parasitologia Veterinária**, v. 25, n. 3, p. 317-326, jul./set. 2016. DOI: <https://doi.org/10.1590/S1984-29612016053>.

LOBO, I. K. C. **Análise do transcriptoma de *Fusarium decemcellulare* agente causal do superbrotamento em guaranazeiro (*Paullinia cupana* var. *Sorbilis*)**. 2016. 94 p. Dissertação (Mestrado em Biotecnologia) — Universidade Federal do Amazonas, Manaus. Orientador: Gilvan Ferreira da Silva; coorientador: Michel Eduardo Bezeza Yamagishi.

MCCOUCH, S. R.; WRIGHT, M. H.; TUNG, C. W.; MARON, L. G.; MCNALLY, K. L.; FITZGERALD, M.; SINGH, N.; DECLERCK, G.; AGOSTO-PEREZ, F.; KORNILIEV, P.; GREENBERG, A. J.; NAREDO, M. E. B.; MERCADO, S. M. Q.; HARRINGTON, S. E.; SHI, Y.; BRANCHINI, D. A.; KUSER-FALCÃO, P. R.; LEUNG, H.; EBANA, K.; YANO, M.; EIZENGA, G.; MCCLUNG, A.; MEZEY, J. Open access resources for genome-wide association mapping in rice. **Nature Communications**, v. 7, 10532, 2016. DOI: <https://doi.org/10.1038/ncomms10532>.

SILVA, J. M. da; GIACHETTO, P. F.; SILVA, L. O. da; CINTRA, L. C.; PAIVA, S. R.; YAMAGISHI, M. E. B.; CAETANO, A. R. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. **BMC Genomics**, v. 17, 454, 2016. DOI: <https://doi.org/10.1186/s12864-016-2752-9>.

ZERLOTINI NETO, A.; CINTRA, L. C. **Análise de dados de RNA-Seq utilizando o Galaxy**. Campinas: Embrapa Informática Agropecuária, 2016. 36 p. il. (Embrapa Informática Agropecuária. Documentos, 149). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1064217>. Acesso em: 9 ago. 2024.

2017

BELESINI, A. A.; CARVALHO, F. M. S.; TELLES, B. R.; CASTRO, G. M. de; GIACHETTO, P. F.; VANTINI, J. S.; CARLIN, S. D.; CAZETTA, J. O.; PINHEIRO, D. G.; FERRO, M. I. T. De novo transcriptome assembly of sugarcane leaves submitted to prolonged water-deficit stress. **Genetics and Molecular Research**, v. 16, n. 2, gmr16028845, 2017. DOI: <http://dx.doi.org/10.4238/gmr16028845>.

BUZANSKAS, M. E.; GROSSI, D. do A.; VENTURA, R. V.; SCHENKEL, F. S.; CHUD, T. C. S.; STAFUZZA, N. B.; ROLA, L. D.; MEIRELLES, S. L. C.; MOKRY, F. B.; MUDADU, M. de A.;

HIGA, R. H.; SILVA, M. V. G. B. da; ALENCAR, M. M. de; REGITANO, L. C. de A; MUNARI, D. P. Candidate genes for male and female reproductive traits in Canchim beef cattle. **Journal of Animal Science and Biotechnology**, v. 8, 67, 2017. DOI: <https://doi.org/10.1186/s40104-017-0199-8>.

BUZANSKAS, M. E.; VENTURA, R. V.; CHUD, T. C. S.; BERNARDES, P. A.; SANTOS, D. J. de A.; REGITANO, L. C. de A.; ALENCAR, M. M. de; MUDADU, M. de A.; ZANELLA, R.; SILVA, M. V. G. B. da; LI, C.; SCHENKEL, F. S.; MUNARI, D. P. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. **PLoS One**, v. 12, n. 2, e0171660, 2017. DOI: <https://doi.org/10.1371/journal.pone.0171660>.

COELHO, W. D. **Soluções computacionais para a coleta e análise de dados em programas de melhoramento genético animal**. 2017. 75 p. Dissertação (Mestrado em Zootecnia) — Universidade Federal de Pelotas, Pelotas.

GOUVEIA, J. J. de S.; PAIVA, S. R.; MCMANUS, C. M.; CAETANO, A. R.; KIJAS, J. W.; FACÓ, O.; AZEVEDO, H. C.; ARAUJO, A. M. de; SOUZA, C. J. H. de; YAMAGISHI, M. E. B.; CARNEIRO, P. L. S.; LÔBO, R. N. B.; OLIVEIRA, S. M. P. de; SILVA, M. V. G. B. da. Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. **Livestock Science**, n. 197, p. 36-45, 2017. DOI: <https://doi.org/10.1016/j.livsci.2017.01.006>.

STAFUZZA, N. B.; ZERLOTINI, A.; LOBO, F. P.; YAMAGISHI, M. E. B.; BUZANSKAS, M. E.; CHUD, T. C. S.; Caetano, A. R.; MUNARI, D. P.; GARRICK, D. J.; MACHADO, M. A.; MARTINS, M. F.; CARVALHO, M. R.; COLE, J. B.; SILVA, M. V. G. B. da. Genetic variants with potential loss of function in Gyr, Girolando, and Guzerat cattle breeds by resequencing. **Journal of Animal Science**, v. 95, p. 81, 2017. Suplemento 4, Resumo 165. Edição de Abstracts do ASAS-CSAS Annual Meeting and Trade Show, Baltimore, 2017. DOI: <http://dx.doi.org/10.2527/asasann.2017.165>.

STAFUZZA, N. B.; ZERLOTINI, A.; LOBO, F. P.; YAMAGISHI, M. E. B.; CHUD, T. C. S.; CAETANO, A. R.; MUNARI, D. P.; GARRICK, D. J.; MACHADO, M. A.; MARTINS, M. F.; CARVALHO, M. R.; COLE, J. B.; SILVA, M. V. G. B. da. Single nucleotide variants and InDels identified from whole-genome re-sequencing of Guzerat, Gyr, Girolando and Holstein cattle breeds. **PLoS One**, v. 12, n. 3, e0173954, 2017. DOI: <https://doi.org/10.1371/journal.pone.0173954>.

KASARAPU, P.; PORTO-NETO, L. R.; FORTES, M. R. S.; LEHNERT, S. A.; MUDADU, M. A.; COUTINHO, L.; REGITANO, L.; GEORGE, A.; REVERTER, A. The *Bos taurus*-*Bos indicus* balance in fertility and milk related genes. **PLoS One**, v. 12, n. 8, e0181930, 2017. DOI: <https://doi.org/10.1371/journal.pone.0181930>.

MORO, C. F.; GASPAS, M.; SILVA, F. R. da; PATTATHIL, S.; HAHN, M. G.; SALGADO, I.; BRAGA, M. R. S-nitrosoglutathione promotes cell wall remodelling, alters the transcriptional profile and induces root hair formation in the hairless root hair defective 6 (rhd6) mutant of *Arabidopsis thaliana*. **New Phytologist**, v. 213, n. 4, p. 1771-1786, Mar. 2017. DOI: <https://doi.org/10.1111/nph.14309>.

SOMAVILLA, A. L.; REGITANO, L. C. de A.; ROSA, G. J. M.; MOKRY, F. B.; MUDADU, M. de A.; TIZIOTO, P. C.; OLIVEIRA, P. S. N. de; SOUZA, M. M. de; COUTINHO, L. L.; MUNARI, D. P. Genome-enabled prediction of breeding values for feedlot average daily weight gain in Nelore cattle. *G3: Genes, Genomes, Genetics*, v. 7, n. 6, p. 1855-1859, 2017. DOI: <https://doi.org/10.1534/g3.117.041442>.

VILLELA, L. C. V.; ALVES, A. L.; VARELA, E. S.; YAMAGISHI, M. E. B., GIACHETTO, P. F.; SILVA, N. M. A. da; PONZETTO, J. M.; PAIVA, S. R.; CAETANO, A. R. Complete mitochondrial genome from South American catfish *Pseudoplatystoma reticulatum* (Eigenmann & Eigenmann) and its impact in Siluriformes phylogenetic tree. *Genetica*, v. 145, n. 1, p. 51-66, Feb. 2017. DOI: <https://doi.org/10.1007/s10709-016-9945-7>.

YAMAGISHI, M. E. B. Expanding the grammar of Biology. In: YAMAGISHI, M. E. B. **Mathematical grammar of Biology**. Switzerland: Springer, 2017. chap. 2, p. 29-54. (Springer briefs in Mathematics). DOI: https://doi.org/10.1007/978-3-319-62689-5_3.

YAMAGISHI, M. E. B. "In God we trust; all others, bring data". In: YAMAGISHI, M. E. B. **Mathematical grammar of Biology**. Switzerland: Springer, 2017. chap. 4, p. 55-78. (Springer briefs in Mathematics). DOI: https://doi.org/10.1007/978-3-319-62689-5_4.

YAMAGISHI, M. E. B. Modeling human nucleotide frequencies. In: YAMAGISHI, M. E. B. **Mathematical grammar of Biology**. Switzerland: Springer, 2017. chap. 2, p. 9-28. (Springer briefs in Mathematics). DOI: https://doi.org/10.1007/978-3-319-62689-5_2.

2018

ANDREOTTI, R.; GIACHETTO, P. F.; CUNHA, R. C. Advances in tick vaccinology in Brazil: from gene expression to immunoprotection. *Frontiers In Bioscience*, v. 10, p. 142, 2018. DOI: <https://doi.org/10.3389/fnmol.2018.00277>.

CESAR, A. S. M.; REGITANO, L. C. A.; REECY, J. M.; POLETI, M. D.; OLIVEIRA, P. S. N. de; OLIVEIRA, G. B. de; MOREIRA, G. C. M.; MUDADU, M. A.; TIZIOTO, P. C.; KOLTES, J. E.; FRITZ-WATERS, E.; KRAMER, L.; GARRICK, D.; BEIKI, H.; GEISTLINGER, L.; MOURÃO, G. B.; ZERLOTINI, A.; COUTINHO, L. L. Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. *BMC Genomics*, v. 19, 499, 2018. DOI: <https://doi.org/10.1186/s12864-018-4871-y>.

DINIZ, M. R. V.; PAIVA, A. L. B.; GUERRA-DUARTE, C.; NISHIYAMA JÚNIOR, M. Y.; MUDADU, M. A.; OLIVEIRA, U. de; BORGES, M. H.; YATES, J. R.; JUNQUEIRA-DE-AZEVEDO, I. de. An overview of Phoneutria nigriventer venom using combined transcriptomic and proteomic approaches. *PLoS One*, v. 13, n. 8, e0200628, 2018. DOI: <https://doi.org/10.1371/journal.pone.0200628>.

GUIMARÃES, M. Z. P.; VECCHI, R.; VITÓRIA, G.; SOCHACKI, J. K.; PAULSEN, B. S.; LIMA, I.; SILVA, F. R. da; COSTA, R. F. M. DA; CASTRO, N. G.; BRETON, L.; REHEN, S. K. Generation of iPSC-derived human peripheral sensory neurons releasing substance P elicited by TRPV1 agonists. *Frontiers in Molecular Neuroscience*, v. 11, 277, Aug. 2018.

NASCIMENTO, G. B. do. **Estratégias de imputação e associação genômica com dados de sequenciamento para características de produção de leite na raça Gir**. 2018. 60 p. Tese (Doutorado em Genética e Melhoramento Animal) — Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio Mesquita Filho", Campus de Jaboticabal, Jaboticabal.

PEREIRA, C. M.; CARVALHO, A. C. de; SILVA, F. R. da; MELENDEZ, M. E.; LESSA, R. C.; ANDRADE, V. C. C.; KOWALSKI, L. P.; VETTORE, A. L.; CARVALHO, A. L. In vitro and in silico validation of CA3 and FHL1 downregulation in oral cancer. **BMC Cancer**, v. 18, 193, 2018. DOI: <https://doi.org/10.1186/s12885-018-4077-3>.

SUÁREZ, W. A. B.; VANTINI, J. da S.; DUDA, R. M.; GIACHETTO, P. F.; CINTRA, L. C.; FERRO, M. I. T.; OLIVEIRA, R. A. de. Predominance of syntrophic bacteria, Methanosaeta and Methanoculleus in a two-stage up-flow anaerobic sludge blanket reactor treating coffee processing wastewater at high organic loading rate. **Bioresour Technol**, v. 268, p. 158-168, Nov. 2018. DOI: <https://doi.org/10.1016/j.biortech.2018.06.091>.

TELLES, B. R.; CARVALHO, F. M. de S.; VANTINI, J. da S.; BELESINI, A. A.; CASTRO, G. M. de; GIACHETTO, P. F.; CARLIN, S. D.; SILVA, T. R. da; PINHEIRO, D. G.; CAZETTA, J. O.; FERRO, M. I. T. Prolonged water deficit reveals new profile of sugarcane gene expression and metabolic pathway related to tolerance. **Sugar Tech**, v. 21, n. 3, p. 451-461, May-June 2019. DOI: <https://doi.org/10.1007/s12355-018-0674-3>.

2019

AZEVEDO, D. A. A. de; MONTEIRO, J. P.; PINHEIRO, R. R.; MUDADU, M. de A.; ANDRIOLI, A.; ARAÚJO, J. F.; SOUSA, A. L. M. de; SIDER, L. H.; PEIXOTO, R. M.; TEIXEIRA, M. F. da S. Molecular characterization of circulating strains of small ruminant lentiviruses in Brazil based on complete gag and pol genes. **Small Ruminant Research**, v. 177, p. 160-166, 2019. DOI: <https://doi.org/10.1016/j.smallrumres.2019.06.011>.

DINIZ, W. J. da S. **Identificação de redes gênicas de coexpressão e dos mecanismos regulatórios associados à composição mineral e qualidade de carne em bovinos**. 2019. 221 p. Tese (Doutorado em Filosofia) – Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos, São Carlos.

MORÉ, D. D.; CARDOSO, F. F.; MUDADU, M. de A.; MALAGÓ JÚNIOR, W.; GULIAS-GOMES, C. C.; SOLLERO, B. P.; IBELLI, A. M. G.; COUTINHO, L. L.; REGITANO, L. C. A. Network analysis uncovers putative genes affecting resistance to tick infestation in Braford cattle skin. **BMC Genomics**, v. 20, 998, 2019. DOI: <https://doi.org/10.1186/s12864-019-6360-3>.

PAIM, T. do P.; MCMANUS, C.; VIEIRA, F. D.; OLIVEIRA, S. R. de M.; FACÓ, O.; AZEVEDO, H. C.; ARAÚJO, A. M. de; MORAES, J. C. F.; YAMAGISHI, M. E. B.; CARNEIRO, P. L. S.; CAETANO, A. R.; PAIVA, S. R. Validation of a customized subset of SNPs for sheep breed assignment in Brazil. **Pesquisa Agropecuária Brasileira**, v. 54, e00506, 2019. DOI: <https://doi.org/10.1590/S1678-3921.pab2019.v54.00506>.

PAIVA, A. L. B.; MUDADU, M. de A.; PEREIRA, E. H. T.; MARRI, C. A.; GUERRA-DUARTE, C.; DINIZ, M. R. V. Transcriptome analysis of the spider *Phoneutria pertyi* venom glands reveals

novel venom components for the genus *Phoneutria*. **Toxicon**, v. 163, p. 59-69, May 2019. DOI: <https://doi.org/10.1016/j.toxicon.2019.03.014>.

PEIXOTO, J. de O.; SAVOLDI, I. R.; IBELLI, A. M. G.; CANTÃO, M. E.; JAENISCH, F. R. F.; GIACHETTO, P. F.; SETTLES, M. L.; ZANELLA, R.; MARCHESI, J. A. P.; PANDOLFI, J. R.; COUTINHO, L. L.; LEDUR, M. C. Proximal femoral head transcriptome reveals novel candidate genes related to epiphyseolysis in broiler chickens. **BMC Genomics**, v. 20, 1031, 2019. DOI: <https://doi.org/10.1186/s12864-019-6411-9>.

SOUSA, E. M. R.; SUZART, C.; COSTA, S. N.; COSTA, M. G. C.; ALMEIDA, A. A. F. de; COELHO FILHO, M. A.; SILVA, F. R. da; SOARES FILHO, W. dos S.; MICHELI, F.; GESTEIRA, A. da S. Transcriptomic analysis related to the flowering of the citrus hybrid *Microcitrangemonia*. **Current Plant Biology**, v. 18, 100097, May 2019. DOI: <https://doi.org/10.1016/j.cpb.2018.12.003>.

VALENTE, J. B.; PEREIRA, F. S.; STEMPKOWSKI, L. A.; FARIAS, M.; KUHNEM, P.; LAU, D.; FAJARDO, T. V. M.; NHANI JÚNIOR, A.; CASA, R. T.; BOGO, A.; SILVA, F. N. da. A novel putative member of the family Benyviridae is associated with soilborne wheat mosaic disease in Brazil. **Plant Pathology**, v. 68, n. 3, p. 588-600, Apr. 2019. DOI: <https://doi.org/10.1111/ppa.12970>.

2020

ANGELO, P. C. da S.; YAMAGISHI, M. E. B.; CRUZ, J. C. da; SILVA, G. F. da; GASPAROTTO, L. Differential expression and structural polymorphism in rubber tree genes related to South American leaf blight resistance. **Physiological and Molecular Plant Pathology**, v. 110, 101477, Apr. 2020. DOI: <https://doi.org/10.1016/j.pmpp.2020.101477>.

DINIZ, W. J. da S.; BANERJEE, P.; MAZZONI, G.; COUTINHO, L. L.; CESAR, A. S. M.; AFONSO, J.; GROMBONI, C. F.; NOGUEIRA, A. R. A.; KADARMIDEEN, H. N.; REGITANO, L. C. de A. Interplay among miR-29 family, mineral metabolism, and gene regulation in *Bos indicus* muscle. **Molecular Genetics and Genomics**, v. 295, p. 1113-1127, 2020. DOI: <https://doi.org/10.1007/s00438-020-01683-9>.

GIACHETTO, P. F.; CUNHA, R. C.; NHANI JÚNIOR, A.; GARCIA, M. V.; FERRO, J. A.; ANDREOTTI, R. Gene expression in the salivary gland of *Rhipicephalus (Boophilus)* microplus fed on tick-susceptible and tick-resistant hosts. **Frontiers in Cellular and Infection Microbiology**, v. 9, 477, Jan. 2020. DOI: <https://doi.org/10.3389/fcimb.2019.0047>.

LEMOS, M. P. L.; SARAIVA, M. M. S.; LEITE, E. L.; SILVA, N. M. V.; VASCONCELOS, P. C.; GIACHETTO, P. F.; FREITAS NETO, O. C.; GIVISIEZ, P. E. N.; GEBREYES, W. A.; OLIVEIRA, C. J. B. The posthatch prophylactic use of ceftiofur affects the cecal microbiota similar to the dietary sanguinarine supplementation in broilers. **Poultry Science**, v. 99, n. 11, p. 6013-6021, Nov. 2020. DOI: <https://doi.org/10.1016/j.psj.2020.06.078>.

LOBO, I. K. C.; NASCIMENTO, A. R. do; YAMAGISHI, M. E. B.; GUIGUEN, Y.; SILVA, G. F. da; SEVERAC, D.; AMARAL, A. da C.; REIS, V. R.; O' SULLIVAN, F. L. A. Transcriptome of tambaqui *Colossoma macropomum* during gonad differentiation: different molecular signals

leading to sex identity. **Genomics**, v. 112, n. 3, p. 2478-2488, May 2020. DOI: <https://doi.org/10.1016/j.ygeno.2020.01.022>.

MUDADU, M. de A.; ZERLOTINI, A. Machado: open source genomics data integration framework. **GigaScience**, v. 9, n. 9, g1aa097, 2020. DOI: <https://doi.org/10.1093/gigascience/g1aa097>.

QUEIROZ, C. A. de; MATOS, K. da S.; LOBO, I. K. C.; SOUSA, S. B. de; MUNIZ, C. R.; YAMAGISHI, M. E. B.; ZOCCOLO, G. J.; HANADA, R. E.; GASPAROTTO, L.; SOUSA, N. R.; ANDRADE, E. V. de; CORDEIRO, I. B.; MELO, A. A. M. A.; SILVA, G. F. da Morpho-anatomical and molecular characterization of the oversprouting symptoms caused by *Fusarium decemcellulare* in guarana plants (*Paullinia cupana* var. *sorbilis*). **Tropical Plant Biology**, v. 13, p. 247-286, 2020. DOI: <https://doi.org/10.1007/s12042-020-09256-1>.

SOUZA, M. M. de; ZERLOTINI, A.; ROCHA, M. I. P.; BRUSCADIN, J. J.; DINIZ, W. J. da S.; CARDOSO, T. F.; CESAR, A. S. M.; AFONSO, J.; ANDRADE, B. G. N.; MUDADU, M. de A.; MOKRY, F. B.; TIZIOTO, P. C.; OLIVEIRA, P. S. N. de; NICIURA, S. C. M.; COUTINHO, L. L.; REGITANO, L. C. de A. Allele-specific expression is widespread in *Bos indicus* muscle and affects meat quality candidate genes. **Scientific Reports**, v. 10, n. 1, 10204, 2020. DOI: <https://doi.org/10.1038/S41598-020-67089-0>.

2021

BRUSCADIN, J. J.; SOUZA, M. M. de; OLIVEIRA, K. S. de; ROCHA, M. I. P.; AFONSO, J.; CARDOSO, T. F.; ZERLOTINI NETO, A.; COUTINHO, L. L.; NICIURA, S. C. M.; REGITANO, L. C. de A. Muscle allele-specific expression QTLs may affect meat quality traits in *Bos indicus*. **Scientific Reports**, v. 11, 7321, 2021. DOI: <https://doi.org/10.1038/s41598-021-86782-2>.

CARDOSO, T. F.; COUTINHO, L. L.; BRUSCADIN, J. J.; DINIZ, W. J. da S.; PETRINI, J.; ANDRADE, B. G. N.; OLIVEIRA, P. S. N. de; POLETI, M. D.; CESAR, A. S. M.; SILVEIRA, J. C. da; CHIARATTI, M. R.; ZERLOTINI, A.; MOURÃO, G. B.; REGITANO, L. C. de A. Multi-omics approach reveals miR-SNPs affecting muscle fatty acids profile in Nelore cattle. **Genes**, v. 12, n. 1, 67, Jan. 2021. DOI: <https://doi.org/10.3390/genes12010067>.

FERREIRA, J. R.; TORRES, G. A. M.; CONSOLI, L.; ANDRIOLIO, G.; SCAGLIUSI, S. M. M.; NHANI JUNIOR, A.; TURCHETTO, C.; DEUNER, C. C.; GODDARD, R.; NICHOLSON, P. Quantitative trait loci conferring blast resistance in hexaploid wheat at adult plant stage. **Plant Pathology**, v. 70, n. 1, p. 100-109, Jan. 2021. DOI: <https://doi.org/10.1111/ppa.13278>

GUIMARÃES, P. S.; SCHENK, J. C. M.; CINTRA, L. C.; GIACHETTO, P. F.; SILVAROLLA, M. B.; PADILHA, L.; MALUF, M. P. Large-scale prospection of genes on caffeine-free *Coffea arabica* plants — discovery of novel markers associated with development and secondary metabolism. **Plant Gene**, v. 27, 100314, Sept. 2021. DOI: <https://doi.org/10.1016/j.plgene.2021.100314>.

HUL, L. M.; IBELLI, A. M. G.; SAVOLDI, I. R.; MARCELINO, D. E. P.; FERNANDES, L. T.; PEIXOTO, J. O.; CANTÃO, M. E.; HIGA, R. H.; GIACHETTO, P. F.; COUTINHO, L. L.; LEDUR,

M. C. Differentially expressed genes in the femur cartilage transcriptome clarify the understanding of femoral head separation in chickens. **Scientific Reports**, v. 11, n. 1, 17965, 2021. DOI: <https://doi.org/10.1038/s41598-021-97306-3>.

LYRA, M. do C. C. P. de; TAKETANI, R. G.; FREITAS, A. D. S. de; SILVA, C. E. R. S. e; MERGULHÃO, A. C. do E. S.; SILVA, M. L. R. B. da; ANTUNES, J. E. L. S.; ARAÚJO, A. S. F. de; GIACHETTO, P. F. Structure and diversity of bacterial community in semiarid soils cultivated with prickly-pear cactus (*Opuntia ficus-indica* (L.) Mill.). **Anais da Academia Brasileira de Ciências**, v. 93, n. 3, e20190183-9, 2021. DOI: <https://doi.org/10.1590/0001-376520212019018>.

PAIM, T. P.; PAIVA, S. R.; TOLEDO, N. M. de; YAMAGISHI, M. B.; CARNEIRO, P. L. S.; FACÓ, O.; ARAUJO, A. M. de; AZEVEDO, H. C.; CAETANO, A. R.; BRAGA, R. M.; MCMANUS, C. Origin and population structure of Brazilian hair sheep breeds. **Animal Genetics**, v. 52, n. 4, p. 492-504, Aug. 2021. DOI: <https://doi.org/10.1111/age.13093>.

SAVI, A.; GORAYEB, E. S.; NASCIMENTO, S. C.; FARIA, C. B. de; FAJARDO, T. V. M.; NHANI JÚNIOR, A.; BOGO, A.; CASA, R. T.; SILVA, F. N. da. Near-complete genome sequence and seed transmission evaluation of *Physalis rugose* mosaic virus from southern Brazil. **Ciência Rural**, v. 51, n. 4, e20200702, 2021. DOI: <https://doi.org/10.1590/0103-8478cr20200702>.

SOUZA, D. D. de; QUEIROZ, A. P. de; PEREIRA, F. S.; DIANESE, E. de C.; FAJARDO, T. V. M.; NHANI JÚNIOR, A.; LAU, D.; SILVA, L. A. da; RIBEIRO, B. M.; COELHO, A. S. G.; AGUIAR, R. W. de S.; MELLO, R. N. de; SILVA, F. N. da. Molecular characterization and sequence analysis of four Brazilian rice stripe necrosis virus isolates. **Archives of Virology**, v. 166, n. 6, p. 1763-1767, June 2021. DOI: <https://doi.org/10.1007/s00705-021-05037-7>.

VIANA, M. J. A.; ZERLOTINI, A.; MUDADU, M. de A. Plant co-expression annotation resource: a web server for identifying targets for genetically modified crop breeding pipelines. **BMC Bioinformatics**, v. 22, 46, 2021. DOI: <https://doi.org/110.1186/s12859-020-03792-z>.

YAMAGISHI, M. E. B. Is our immune system a powerful vaccine factory? **Genetics And Molecular Biology**, v. 44, n. 1, e20200468, 2021. Supplement 1. Letter to the editor. DOI: <https://doi.org/10.1590/1678-4685-GMB-2020-0468>.

2022

RUSSO, G. H. **Identificação de polimorfismos de nucleotídeo único a partir do sequenciamento do genoma completo de touros da raça Gir**. 2022. 14 p. Monografia (Bacharelado em Ciências Biológicas) — Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio Mesquita Filho”, Campus de Jaboticabal, Jaboticabal.

SILVA, N. M. L.; IANELLA, P.; YAMAGISHI, M. E. B.; ROCHA, J. L.; TEIXEIRA, A. K.; FARIAS, F. G.; GUERRELHAS, A. C.; CAETANO, A. R. Development and validation of a low-density SNP panel for paternity and kinship analysis and evaluation of genetic variability and structure of commercial Pacific white shrimp (*Litopenaeus vannamei*) populations from Brazil. **Aquaculture**, v. 560, 738540, Nov. 2022. DOI: <https://doi.org/10.1016/j.aquaculture.2022.738540>.

Publicações na Imprensa

HEBERLÊ, D. **Ferramenta genética faz análise de linhagens de camarão cinza**. 2021. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/60909107/ferramenta-genetica-faz-analise-de-linhagens-de-camarao-cinza>. Acesso em: 9 ago. 2024.

SOUZA, S. **Pesquisa identifica genes envolvidos na resistência ao mal-das-folhas da seringueira**. 2021. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/60012973/pesquisa-identifica-genes-envolvidos-na-resistencia-ao-mal-das-folhas-da-seringueira>. Acesso em: 9 ago. 2024.

HEBERLÊ, M. D. **Produção de peixes: cresce procura pelo teste TambaPlus**. 2020. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/54666474/producao-de-peixes-cresce-procura-pelo-teste-tambaplus>. Acesso em: 9 ago. 2024.

SOUZA, S. **Novos microorganismos de interesse econômico são encontrados em rios amazônicos**. 2020. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/54173860/novos-microorganismos-de-interesse-economico-sao-encontrados-em-rios-amazonicos>. Acesso em: 9 ago. 2024.

YAMAGISHI, M. E. B. **A gramática matemática da Biologia**. 2017. Disponível em: <https://indicadoreconomico.com.br/coluna-2/coluna-2-b/artigo-a-gramatica-matematica-da-biologia/>. Acesso em: 9 ago. 2024.

PESQUISA genômica da Embrapa pode aumentar produtividade do gado nelore. **Globo Rural**, 14 jun. 2016. Disponível em: <http://revistagloborural.globo.com/Noticias/Criacao/Boi/noticia/2016/06/pesquisa-genomica-da-embrapa-pode-aumentar-produtividade-do-gado-nelore.html>. Acesso em: 9 ago. 2024.

Do gado Nelore ao genoma humano. **Jornal da Unicamp**, ano 29, n. 638, p. 1, 21 a 27 set. 2015.

ORSI, C. Análise de DNA de gado abre caminho para detecção de variações no genoma humano. **Jornal da Unicamp**, ano 2015, n. 638, p. 3, 21 a 27 set. 2015. Disponível em: <https://unicamp.br/unicamp/ju/638/analise-de-dna-de-gado-abre-caminho-para-deteccao-de-variacoes-no-genoma-humano>. Acesso em: 9 ago. 2024.

SUGIMOTO, L. Modelos podem subsidiar certificação de ovinos. **Jornal da Unicamp**, ano 2014, n. 610, p. 5, 13 a 19 ou. 2014. Disponível em: https://www.unicamp.br/unicamp/sites/default/files/jornal/paginas/ju_610_paginacor_05_web.pdf. Acesso em: 9 ago. 2024.

Capítulos de Livros

GIACHETTO, P. F.; HIGA, R. H. Bioinformatics applied to agriculture. In: MASSRUHÁ, S. M. F. S.; LEITE, M. A. de A.; LUCHIARI JUNIOR, A.; ROMANI, L. A. S. (ed.). **Information and**

communication technologies and their relations with agriculture. Brasília, DF: Embrapa, 2016. ch. 4, p. 67-86.

GIACHETTO, P. F. Transcriptoma do carrapato dos bovinos. In: ANDREOTTI, R.; GARCIA, M. V.; KOLLER, W. W. (ed.). **Carrapatos na cadeia produtiva de bovinos.** Brasília, DF: Embrapa, 2019. cap. 16, p. 205-222. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1107162>. Acesso em: 9 ago. 2024.

ZERLOTINI NETO, A.; NHANI JÚNIOR, A.; VIEIRA, F. D.; CINTRA, L. C.; MUDADU, M. de A.; FALCÃO, P. R. K.; GIACHETTO, P. F. Aplicações da bioinformática na agricultura. In: MASSRUHÁ, S. M. F. S.; LEITE, M. A. de A.; OLIVEIRA, S. R. de M.; MEIRA, C. A. A.; LUCHIARI JUNIOR, A.; BOLFE, E. L. (ed.). **Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas.** Brasília, DF, Embrapa, 2020. cap. 10, p. 234-257. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1126268>. Acesso em: 9 ago. 2024.

Livro

TORRES, T. Z.; VISOLI, M. C.; FALCÃO, P. R. K.; SOUZA, M. I. F.; GIACHETTO, P. F.; NHANI JUNIOR, A.; CINTRA, L. C.; CUNHA, L. M. S.; BARBOSA, L. A. F. **Gestão de dados de pesquisa no laboratório multiusuário de bioinformática da Embrapa.** Ponta Grossa: Atena, 2022. 57 p. DOI: <https://doi.org/10.22533/at.ed.837220802>.

Patentes

HONGO, J. A.; LOBO, F. P. **POsitive selecTION – POTION.** Titular: Empresa Brasileira de Pesquisa Agropecuária. BR512015000840-4. Depósito: 13 mar. 2015. Concessão: 8 mar. 2016.

Programas de Computador

HIGA, R. H.; MUDADU, M. de A.; TIZIOTO, P. C.; REGITANO, L. C. de A. **Rpaternity.** Versão 2.0. Campinas: Embrapa Informática Agropecuária, 2012. 1 CD-ROM.

HONGO, J. A. **POTION.** Versão 1.0.1. Campinas: Embrapa Informática Agropecuária, 2014. 1 CD-ROM.

MUDADU, M.A.; ZERLOTINI NETO, A. **MACHADO:** um framework para armazenar, buscar e visualizar informação biológica. Versão 0.5.0. Campinas: Embrapa Informática Agropecuária, 2018. Disponível em: <http://www.machado.cnptia.embrapa.br>. Acesso em: 14 ago. 2024.

MUDADU, M.A. **PlantAnnot.** Campinas: Embrapa Informática Agropecuária, 2021. Acessado por meio da Plataforma AgroAPI da Embrapa. Mais informações, disponíveis em: <https://www.agroapi.cnptia.embrapa.br/portal/assets/docs/plantannot.pdf>. Acesso em: 14 ago. 2024.

VELLOZO, A. F.; YAMAGISHI, M. E. B. **FIOMA**. Versão 1.0. Campinas: Embrapa Informática Agropecuária, 2014. 1 CD-ROM.

YAMAGISHI, M. E. B.; LOBO, F. P.; VELLOZO, A. F. **Blast.pm**. Versão 0.01 Campinas: Embrapa Informática Agropecuária, 2012. 1 CD-ROM.

Anexo 3 — Formação de recursos humanos

Alunos de Pós-Graduação do LMB

Mestrado

1. Isabel Pereira Caminha.

CAMINHA, I. P. **Estudos estruturais das enzimas envolvidas com o mecanismo de produção de ácidos orgânicos da bactéria gluconacetobacter diazotrophicus utilizando métodos computacionais**. 2011. 80 p. Dissertação (Mestrado em Genética e Biologia Molecular) – Instituto de Biologia, Universidade Estadual de Campinas, Campinas. Orientador: Paula Regina Kuser Falcão. DOI: <https://doi.org/10.47749/T/UNICAMP.2011.795672>.

2. Lucas Miguel de Carvalho.

CARVALHO, L. M. de. **Avaliação de estratégias para análise 'de novo' de transcriptomas**. 2013. 123 p. Dissertação (Mestrado em Ciência da Computação) – Instituto de Computação, Universidade Estadual de Campinas, Campinas. Orientador: Zaroni Dias, coorientador: Felipe Rodrigues da Silva. DOI: <https://doi.org/10.47749/T/UNICAMP.2015.965252>.

3. Luiz Augusto Eijy Nagai.

NAGAI, L. A. E. **Identificação de genes relacionados à maciez da carne de bovinos das raças Angus e Nelore por análise da expressão diferencial**. 2015. 64 p. Dissertação (Mestrado em Genética e Biologia Molecular) – Instituto de Biologia, Universidade Estadual de Campinas, Campinas. Orientador: Paula Regina Kuser Falcão, coorientador: Adhemar Zerlotini Neto. DOI: <https://doi.org/10.47749/T/UNICAMP.2015.950670>.

4. Giovanni Marques de Castro.

CASTRO, G. M. de. **Identificação da quasispecies Papaya ringspot vírus em uma biblioteca de cDNA de Fevillea cordifolia**. 2015. 47 p. Dissertação (Mestrado em Genética e Biologia Molecular) – Instituto de Biologia, Universidade Estadual de Campinas, Campinas. Orientador: Felipe Rodrigues da Silva, coorientador: Francisco Pereira Lobo. DOI: <https://doi.org/10.47749/T/UNICAMP.2015.950670>.

5. Jorge Augusto Hongo.

HONGO, J. A. **KOMODO 2 — Large-scale genomic inference**. 2015. 99 p. Dissertação (Mestrado em Ciência da Computação) – Instituto de Computação, Universidade Estadual de

Campinas, Empresa Brasileira de Pesquisa Agropecuária. Coorientador: Francisco Pereira Lobo. DOI: <https://doi.org/10.47749/T/UNICAMP.2015.962103>.

6. Marcos José Andrade Viana.

VIANA, M. J. A. **Plant Co-expression Annotation Resource 2.0**: uma ferramenta web para a associação de proteínas de função desconhecida a estresses abióticos em plantas. 2021. 77 p. Dissertação (Mestrado em Bioinformática) – Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte. Orientador: Maurício de Alvarenga Mudadu, coorientadores: Francisco Pereira Lobo, Adhemar Zerlotini Neto.

Doutorado

1. Joaquim Manoel da Silva.

SILVA, J. M. da. **Deteção de variações genômicas em bovinos da raça Nelore usando dados de genotipagem e ressequenciamento**. 2015. 107 p. Tese (Doutorado em Genética e Biologia Molecular) – Instituto de Biologia, Universidade Estadual de Campinas, Campinas. Orientador: Michel Eduardo Beleza Yamagishi, coorientador: Alexandre Rodrigues Caetano. DOI: <https://doi.org/10.47749/T/UNICAMP.2015.960141>.

2. Roberto Hirochi Herai.

HERAI, R. H. **Metodologias de bioinformática para deteção e estudo de sequências repetitivas em loci gênicos de transcritos quiméricos**. 2010. 178 p. Tese (Doutorado em Genética e Biologia Molecular) – Instituto de Biologia, Universidade Estadual de Campinas, Campinas. Orientador: Michel Eduardo Beleza Yamagishi. DOI: <https://doi.org/10.47749/T/UNICAMP.2010.480728>.

3. Rodrigo De Vecchi.

DE VECCHI, R. **Caracterização de atividade funcional e expressão gênica de neurônios sensoriais periféricos humanos gerados a partir de células-tronco de pluripotência induzida para uso em modelos pré-clínicos in vitro**. 2018. 123 p. Tese (Doutorado em Genética e Biologia Molecular) – Instituto de Biologia, Universidade Estadual de Campinas, Campinas. Orientador: Felipe Rodrigues da Silva. DOI: <https://doi.org/10.47749/T/UNICAMP.2018.1090654>.

Pós doutorado

1. Augusto Fernandes Vellozo.

Desenvolvimento do software FIOMA, Software que cria a árvore de sufixo para reads (paired-ends e mate-pairs) das novas tecnologias de sequenciamento. Desenvolvido na Embrapa Informática Agropecuária (atual Embrapa Agricultura Digital) com apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior em 2012 sob orientação de Michel Eduardo Beleza Yamagishi.

Cursos ministrados

- Análise de amplicons 16S.
- Análise de dados de *shotgun* metagenômico.

- Análise de RNA-Seq.
- Exploring variation in animal genomes (EBI).
- Montagem de genomas.
- NG110 – Noções Básicas de Bioinformática – Instituto de Biologia, Unicamp (anualmente de 2012 a 2021).
- Uso da plataforma Galaxy.
- Uso do Gbrowse/Chado.

