

Memorial de Atividades, Títulos e Publicações

Concurso de Professor Doutor em RDIDP no Departamento de
Botânica

Instituto de Biociências
Universidade de São Paulo

Diogo Melo
diogro@gmail.com

Conteúdo

Graduação	3
Ciências Moleculares	3
Ciências Biológicas	5
Mestrado	6
Entre mestrado e doutorado	7
Doutorado	8
Pós-doutorado	11
Laboratório de Evolução de Mamíferos	11
A Pandemia e o Observatório Covid-19 BR	15
Ayroles Lab	15
Projetos em andamento e próximos passos	19
Ensino	21
Teoria evolutiva	23
Estatística na era da computação	23
Atividades administrativas	23
Referências	24

Apresentação

Apresento este memorial atendendo ao Edital 038-2023 publicado pela reitoria da Universidade de São Paulo no Diário Oficial da União de 10 de outubro de 2023, referente ao concurso para o cargo de professor doutor na área de “Biologia de Sistemas – Biologia Integrativa e Preditiva”, em Regime de RDIDP, junto ao Departamento de Botânica do Instituto de Biociências da Universidade de São Paulo.

No que se segue apresento minha trajetória acadêmica, incluindo experiência didática e de pesquisa. Destaco alguns projetos em que trabalhei, e como cada parte da minha trajetória contribuiu para minha formação como pesquisador. Ao final, apresento meu currículo completo e comprovantes para os títulos, cursos, apresentações em congressos, e publicações descritas no texto.

Graduação

Ciências Moleculares

Minha trajetória não começou na biologia, mas sim no Curso de Ciências Moleculares (CCM), um curso voltado para formar pesquisadores em ciências exatas e biológicas. Foi minha mãe, a Profa. Nancy Rebouças, pesquisadora no Instituto de Ciências Biomédicas (ICB) da USP, quem primeiro me apresentou esse curso. Ela o conhecia pois um aluno do CCM a havia procurado e estava fazendo iniciação científica em seu laboratório. A profissão da minha mãe teve uma forte influência no meu interesse pela pesquisa científica, sempre gostei de ouvir sobre o dia a dia dela e visitei seu laboratório algumas vezes. Também me sentia atraído pela liberdade que ela tinha para estudar, e via o gosto com que ela se dedicava ao ensino na universidade. Eu me interessei pelo CCM justamente porque ele me daria a oportunidade de transitar por diversas áreas pelas quais eu me interessava, como biologia, física e matemática, pois eu me interessava pela pesquisa de modo geral, mas ainda não havia decidido por uma área específica. Apesar de ser um curso de bacharelado de quatro anos, o CCM não possui ingresso direto pelo vestibular, e a seleção é feita por uma prova no meio do ano letivo, aberta para alunos de qualquer graduação da USP. Assim, eu entrei no curso de Ciências Biológicas do Instituto de Biociências (IB) da USP com a intenção de ingressar no CCM assim que possível. Escolhi a biologia, pois era um curso que eu ficaria feliz de cursar caso não fosse aprovado na prova de seleção do CCM.

Durante o primeiro semestre na biologia descobri que o pai de uma das minhas melhores amigas do ensino médio, com quem eu já tinha uma boa relação, era um dos fundadores e primeiros professores do CCM, o Prof. Hernan Chaimovich do Instituto de Química. Encontrei o Prof. Hernan em um aniversário da sua filha, e ele me incentivou muito a mudar de curso, pois acreditava que eu tinha o perfil de um aluno do CCM. Apesar desses incentivos e de ter sido aprovado na prova de ingresso, durante meu tempo no IB eu me identifiquei com a biologia e considerei desistir da mudança. Foi um aluno do CCM que conheci na época, e que hoje é um bom amigo, que me convenceu a fazer a matrícula. Assim, após apenas seis meses no curso de biologia, mudei para o curso de ciências moleculares, e hoje vejo que esta foi uma decisão extremamente acertada.

O CCM é um curso especial na sua estrutura. Nos dois primeiros anos, o ciclo básico, os alunos cursam um conjunto exigente de disciplinas obrigatórias que fornece uma base sólida em física, química, matemática, computação e aspectos básicos de biologia, como biologia celular, molecular e bioquímica. Os últimos dois anos do curso, o ciclo avançado, são de total liberdade acadêmica. Cada aluno monta, junto com seu orientador, uma grade curricular voltada para atender as demandas da área em que ele escolheu se especializar, podendo cursar qualquer disciplina da universidade, sem restrição. A exigência acadêmica do ciclo básico e a liberdade do ciclo avançado fazem com que cada turma do CCM seja única, formada por pessoas com especialidades diversas, habilidades e interesses complementares, e que ao mesmo tempo

compartilham uma linguagem em comum, estabelecida no ciclo básico. No CCM eu convivi com físicos de partículas a historiadores da ciência, num ambiente que promove a camaradagem e ajuda mútua num processo de aprendizagem desafiador e intenso. Durante o curso, aprendi muito não só com os professores, mas também com meus colegas, e gosto de pensar que também contribuí para o aprendizado do resto da turma. O curso exigiu muito de nós todos e foi um processo de formação crucial para definir a maneira como lido com as pressões do ambiente acadêmico. O ciclo básico do ciências moleculares é tão exigente que fez com que as cobranças do ciclo avançado e da graduação em biologia (quando decidi terminar também este curso) parecessem relativamente brandas. Graças a isso, pude cursar a segunda metade do CCM e o restante do curso de biologia com tranquilidade suficiente para me dedicar também a atividades de pesquisa.

Minha experiência com pesquisa começou durante o ciclo avançado do CCM, no laboratório da Profa. Vera Bohomoletz Henriques, do Departamento de Física Geral do Instituto de Física (IF) da USP. A Profa. Vera lidera um laboratório que integra a teoria de mecânica estatística, uma generalização da termodinâmica, com processos biológicos no nível celular. Meu interesse pelas questões da biologia e os métodos da física me atraíram para essa linha de pesquisa, e comecei um projeto de iniciação científica explorando modelos computacionais para o estudo de transições de fase em membranas biológicas. A ideia do projeto era recuperar em simulações computacionais comportamentos termodinâmicos observados em experimentos com membranas celulares, e com isso entender quais processos são importantes para definir o comportamento das membranas. Para isso, nós utilizamos modelos simples de interações entre os fosfolipídios que compõem as membranas celulares. Este projeto foi meu primeiro contato com uso de simulações em pesquisa, e apresentei resultados preliminares num pequeno congresso em 2007, o *V Brazilian Meeting on Simulational Physics* em Ouro Preto.

Outra influência importante na minha carreira acadêmica foi o Prof. Antônio Carlos Cassola, que era casado com minha mãe. O interesse pelas membranas celulares, que eram o sistema motivador do meu projeto de iniciação científica, e o contato familiar, me levaram a cursar e a atuar como monitor por vários anos na disciplina de fisiologia de membranas que o Prof. Cassola ministrava para o Curso de Medicina da USP. O convívio cotidiano e na universidade com o Prof. Cassola me imbuíu de um profundo respeito pelo debate científico e pelo fazer científico. Suas aulas eram muito ricas e abordavam temas aparentemente díspares, mas sempre ligados por uma narrativa coesa. Participar da sua disciplina foi bastante importante para mim, pois me mostrou como mesmo um tema central da biologia traz consigo uma série de ligações com outras áreas do conhecimento, como a física ou mesmo a filosofia da ciência. Minha atuação em sua disciplina foi o catalisador de um hábito de estar sempre envolvido com monitorias em diversas disciplinas, mesmo que não diretamente relacionadas com a minha pesquisa. Quando entrei na graduação

em biologia, em 2003, Cassola me presenteou com o livro *What is life?*¹, no qual o físico Erwin Schrödinger, um dos principais nomes da mecânica quântica do começo do século XX, discute os aspectos fundamentais dos seres vivos e as propriedades que uma molécula responsável pela herança genética deveria possuir. Hoje em dia, acho interessante pensar como a minha própria trajetória ecoou essa temática, partindo da física e voltando à biologia. Talvez o contato com esta obra de Schrödinger tenha me ajudado a enxergar a conexão entre essas áreas do conhecimento e as possibilidades que a interação entre elas pode gerar.

O trabalho no Instituto de Física também me colocou em contato com outro grupo de pesquisa no mesmo departamento, o dos Profs. Nestor Caticha, do IF, e Renato Vicente, do Instituto de Matemática e Estatística, no qual alguns colegas de graduação faziam as suas iniciações científicas. Este grupo, de mecânica estatística do processamento de informação, era surpreendentemente multidisciplinar e aplicava métodos modernos de física e estatística em problemas vindos da economia, sociologia, biologia e outras áreas da física. Os encontros semanais do grupo eram muito estimulantes intelectualmente e acabaram me influenciando até mais que meu próprio projeto. Mesmo durante o mestrado continuei frequentando essas reuniões. Nelas apresentei prévias de qualificações e palestras, e o grupo sempre me trouxe ideias novas e um ponto de vista externo à biologia, que contribuíram para enriquecer a minha visão teórica e metodológica dos problemas que abordei durante minha formação. Em particular, foi neste grupo que fui apresentado aos métodos Bayesianos, que vieram a se tornar um dos principais focos de estudo e atuação na minha pesquisa. Este contato precoce com a estatística Bayesiana foi um diferencial na minha trajetória acadêmica, já que o tema raramente é apresentado em cursos introdutórios, sendo tradicionalmente abordado somente após um contato com métodos mais tradicionais. Acredito que ter sido apresentado a ela quando eu ainda carregava menos conceitos e preconceitos de outras metodologias contribuiu para o meu domínio destes métodos.

Ciências Biológicas

Tendo me formado no ciências moleculares e encerrado meu projeto de iniciação com a Profa. Vera Henriques, percebi que, apesar do meu interesse pelos métodos utilizados na física, eu me sentia mais atraído pelos problemas levantados pela biologia. Por isso, decidi retornar à graduação em biologia e aplicar as técnicas que havia aprendido na física a problemas com motivações mais biológicas. De volta ao curso de biologia, conversei com alguns professores do Instituto de Biociências, e optei por me juntar ao grupo do Laboratório de Evolução de Mamíferos (LEM) do Prof. Gabriel Marroig. Lembro que quando conversamos pela primeira vez, por intermédio de alguns bons amigos que estudavam em seu laboratório, o Gabriel me disse que o laboratório estava muito cheio e que seria difícil eu entrar naquele momento. Porém, ao descobrir que eu estava vindo da física e que sabia programar, imediatamente falou para marcarmos

¹Schrodinger, R., & Schrödinger, E. (1944). *What is life?: With mind and matter and autobiographical sketches*. Cambridge University Press (re-editado em 1992).

uma conversa. Foi uma sorte, pois nesse momento começamos uma colaboração frutífera e bem sucedida.

Durante minha iniciação científica no LEM, trabalhei com métodos computacionais para produzir estimativas de matrizes de covariância mais resistentes a problemas de amostragem, e que pudessem ser usadas para reconstruir a história evolutiva de fenótipos complexos de forma mais acurada². Este projeto resultou no primeiro artigo em que participei ativamente na pesquisa, escrita e publicação (Marroig et al., 2012). Foi um processo editorial longo até a publicação³. Após uma rejeição acompanhada de boas revisões numa revista prestigiosa, fizemos uma segunda versão bastante modificada e expandida, que foi aceita, incluindo uma parte adicional feita pelo Dr. Guilherme Garcia, que era mestrando na época. Foi meu primeiro projeto com o Guilherme, e mais tarde nós trabalharíamos juntos em outros métodos computacionais aplicados à evolução. Em retrospecto, eu sabia bem menos estatística naquela época do que acreditava, e certamente teria escrito outro manuscrito hoje, mas os métodos e conclusões continuam relevantes e robustos.

Mestrado

Ao concluir a graduação em biologia, eu não tinha dúvidas de que evolução era a área de pesquisa que mais me atraía. A pesquisa em genética de populações e biologia evolutiva depende de modelos matemáticos bem estabelecidos, que me dariam uma oportunidade de usar as técnicas matemáticas e computacionais que eu havia aprendido na física e no curso de ciências moleculares. Além disso, é uma área com diversas perguntas em aberto e que fornece os mecanismos fundamentais para compreender processos biológicos em todos os níveis de organização.

Pensando nisso, ao longo da minha iniciação eu e o Gabriel amadurecemos um projeto de pesquisa para o meu mestrado totalmente baseado em simulações, que fosse relevante para o restante das linhas de pesquisa do LEM. Meu objetivo com esse projeto era trazer para o laboratório algo que em física é algumas vezes chamado de triângulo de Landau⁴: um triângulo equilátero que nos vértices traz as metodologias complementares de **experimento**, **teoria** e **simulação**. O triângulo ilustra como esses três elementos podem atuar em sinergia para promover o desenvolvimento científico. Uma parte fundamental das linhas de pesquisa no Laboratório de Evolução de Mamíferos (LEM) era catalogar os padrões de associação genética entre traços fenotípicos em diversos grupos de mamíferos, e relacionar essa associação com

²Esse pequeno resumo do meu primeiro trabalho ilustra como a minha caminhada para problemas mais biológicos foi na verdade bem gradual. Foi difícil convencer minha banca de iniciação científica que esse projeto era, de fato, biologia.

³Foi neste trabalho que aprendi a não colocar datas em nomes de arquivos de manuscritos. Tive que mudar duas vezes o ano no nome do arquivo.

⁴O nome faz referência ao físico David P. Landau, que sempre usa o triângulo em suas apresentações. Ele é conhecido pelo algoritmo de Wang-Landau, um método de Monte Carlo utilizado amplamente em mecânica estatística. É um algoritmo interessante, que usa uma técnica contra-intuitiva de solução de problemas que consiste em, quando confrontado com um problema complicado, buscar uma maneira de complicá-lo ainda mais, que eventualmente leva a uma solução.

a diversificação morfológica dos grupos. Nesse contexto, existe uma expectativa teórica bem estabelecida de que as associações genéticas devem afetar a resposta evolutiva de curto prazo, e existem dois resultados empíricos relevantes: (1) os padrões de associação genética são relacionados à função desempenhada pelos diferentes traços; (2) os padrões de associação são bastante estáveis em mamíferos, sendo compartilhados por diversos grupos evolutivamente distantes. Pensando no triângulo de Landau, nós nos perguntamos se poderíamos entender, usando um modelo computacional, como essas associações entre diversos caracteres se estabeleciam e como elas eram mantidas. Para isso, eu desenvolvi um modelo baseado em indivíduos que incluía uma arquitetura genética variável, capaz de responder a processos evolutivos. Utilizando esse modelo, mostramos que seleção direcional é um poderoso fator no estabelecimento de associações genéticas, e que seleção estabilizadora era necessária para manter os padrões a longo prazo. Nosso modelo foi o primeiro capaz de simular cenários evolutivos complexos em múltiplos caracteres utilizando uma base genética variável. Este trabalho foi meu primeiro artigo como primeiro autor, e foi publicado em Melo & Marroig (2015).

Durante o mestrado eu me dediquei também a aprofundar meus estudos em estatística Bayesiana, lendo alguns livros texto que hoje norteiam minha prática de pesquisa, em particular os livros *Probability Theory*⁵, de Edwin T. Jaynes, e *Bayesian Data Analysis*⁶, de Andrew Gelman e colaboradores. Esse estudo prolongado ao longo do mestrado, aliado ao contato com o grupo do Prof. Caticha, me forneceram uma base teórica para desenvolver modelos estatísticos customizados para problemas de diversas áreas da biologia, uma habilidade que rendeu colaborações proveitosas durante o doutorado. Listo alguns dos artigos que publiquei como colaborador, nos quais eu fui o principal responsável pela análise de dados, ou pelo desenvolvimento de um modelo adaptado a um problema específico: (Costa et al., 2016; Fukimoto et al., 2020; Hubbe et al., 2016; Wolf et al., 2015) .

O mestrado também foi quando começou minha relação com a linguagem de programação R, que se tornou minha principal ferramenta de trabalho. No ano de 2011, cursei a disciplina de R oferecida pelos Profs. Paulo Inácio Prado e Alexandre Adalardo do Departamento de Ecologia do IB-USP, e depois disso passei a utilizar essa linguagem quase que exclusivamente para minhas análises e projetos. Fui monitor da disciplina de R praticamente todos os anos desde então. Usando essa ferramenta poderosa, no final do mestrado eu comecei um processo de sistematização das rotinas de análise usadas no LEM, que descrevo abaixo.

Entre mestrado e doutorado

Após o mestrado, durante o final de 2012 e todo o ano de 2013, decidi me dedicar totalmente a esse esforço de padronização e implementação em R das rotinas de análise do LEM, com o objetivo de facilitar a adoção dos métodos do laboratório pelos próximos alunos e pela comunidade científica

⁵Jaynes, E. T. (1996). *Probability theory: the logic of science*. St. Louis, MO: Washington University.

⁶Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

em geral. Este trabalho foi financiado com uma bolsa de treinamento técnico da FAPESP. Em colaboração com o Guilherme, que mencionei anteriormente, e partindo das rotinas que ele havia usado no seu mestrado, eu reimplentei o *workflow* do laboratório em R, criando um conjunto de rotinas modulares, eficientes e de fácil disseminação na comunidade científica. O resultado foi o pacote *EvolQG*⁷, que centraliza boa parte das análises em genética quantitativa aplicada a evolução, e está descrito em Melo et al. (2015). Foi durante esse período que realmente entendi como gerenciar um projeto grande em R e passei pelo processo de criar um conjunto de ferramentas abertas para uso por outros pesquisadores. Hoje o pacote conta com mais de 46.000 *downloads*, sendo amplamente utilizado pela comunidade de genética quantitativa.

Além do trabalho com as rotinas em R, neste período comecei a me envolver com um projeto experimental que o Prof. Gabriel vinha planejando há algum tempo, em colaboração com o Prof. Jason Wolf, da Universidade de Bath no Reino Unido. A ideia do projeto era explorar a arquitetura genética de fenótipos complexos utilizando um cruzamento experimental entre linhagens de camundongos que haviam sido selecionadas para mudanças em suas curvas de crescimento. O projeto era grande e ambicioso, envolvendo um longo período de criação da nova colônia de animais. O plano era partir de seis linhagens fundadoras homozigotas e criar uma colônia que seria mantida por várias gerações de cruzamentos, mantendo populações efetivas acima de 400 indivíduos a cada geração. Na sexta e última geração, a F6, criaríamos uma população de mais de 1600 indivíduos, divididos em 160 famílias, que seriam extensamente fenotipados para diversos traços, tais como curva de crescimento, composição corporal, morfologia craniana, entre outros. Além disso, todos os animais das gerações F5 e F6 seriam genotipados utilizando um microarranjo desenhado especificamente para essa colônia. Com esse rico banco de dados, estaríamos na fronteira da investigação experimental da evolução dos efeitos genéticos em caracteres complexos e poderíamos relacionar diretamente o processo evolutivo com a arquitetura genética destes caracteres. Este projeto me interessou muito, principalmente por que ele me daria a oportunidade de avaliar experimentalmente as previsões feitas pelo meu modelo de simulação no mestrado. Era uma oportunidade de estudar como a arquitetura genética interage com um evento de seleção artificial.

Doutorado

O meu doutorado se inseriu neste projeto maior em colaboração com o Prof. Jason. Eu e o Gabriel ficamos responsáveis pela análise da arquitetura genética dos fenótipos multivariados medidos no experimento, com o objetivo principal de estudar efeitos pleiotrópicos (quando um *locus* afeta a variação de mais de um traço simultaneamente). O problema é que, dado o escopo do projeto, eu já sabia que passaria no mínimo 3 anos sem nenhum resultado, e que mesmo após a obtenção dos primeiros resultados ainda seria necessário um longo processo de coleta de dados⁸. Além disso, boa parte do experimento e da coleta de dados aconteceria na Universidade de Bath, onde

⁷<https://cran.r-project.org/web/packages/evolqg/index.html>

⁸Na prática, a coleta de dados só chegou ao fim em 2020, mais de um ano após o final do meu doutorado.

eu poderia passar no máximo alguns meses de cada vez. Essa distância do dia a dia do projeto e a demora para começarmos a ver resultados foram desafios importantes, que me obrigaram a organizar meu tempo de forma inteligente, para poder atender às demandas do projeto principal e, ao mesmo tempo, desenvolver projetos alternativos que me permitissem concluir o doutorado no prazo de 5 anos. Isso fez com que minha tese tivesse uma estrutura incomum, pois cada capítulo usa um conjunto de dados e métodos diferentes, sendo, porém, ligados por uma mesma temática de estudo da covariação genética.

Logo no início do doutorado, em 2014, visitei o laboratório do Prof. Jason por 3 meses. Foi um período rico, em que participei do início do estabelecimento da colônia de camundongos, trabalhando em colaboração com a pós-doutoranda que seria a principal responsável pela criação dos animais, Dra. Eleanor O'Brien. Durante esse período eu colaborei com o Prof. Jason em um projeto paralelo em que desenvolvemos um modelo Bayesiano para analisar os dados de um experimento feito com a ameba *Dictyostelium discoideum*, um eucarioto unicelular que é uma espécie-modelo de colaboração altruística. Nós exploramos um caso de coexistência entre linhagens de *Dictyostelium*, no qual linhagens “egoístas”, que não apresentam comportamento altruísta, conviviam com linhagens “colaboradoras”, que apresentam comportamento altruísta. Essa situação apresenta um desafio teórico, já que, segundo a teoria, a linhagem egoísta, ao se aproveitar do altruísmo de uma linhagem colaboradora sem incorrer os custos do comportamento altruísta, deveria excluir a linhagem colaboradora competitivamente. Essa exclusão não acontece, e nós mostramos que a coexistência era possível devido à presença de demandas conflitantes entre diferentes fases do ciclo de vida das amebas. Considerando todos os componentes da aptidão, em todas as fases do ciclo de vida e não só durante a interação entre linhagens, tanto as linhagens egoístas quanto as colaboradoras apresentavam aptidões semelhantes. Um exemplo de aplicação de genética quantitativa a um campo incomum, este trabalho foi publicado em Wolf et al. (2015), tendo tido uma recepção positiva, sendo inclusive recomendado pela *Faculty of 1000*⁹.

Eu ainda voltaria para a Universidade de Bath uma segunda vez, durante 5 meses em 2016, agora a tempo de participar dos estágios finais da colônia de camundongos. Desta vez participei ativamente da coleta de dados de composição corporal dos animais de F6, além de me tornar o principal responsável por desenhar o microarranjo que depois foi usado para aferir os genótipos de todos os animais das gerações F5 e F6. Foi meu primeiro contato com dados genômicos e ferramentas de bioinformática e novamente um momento de grande aprendizado. Durante essa segunda visita, eu e Jason nos dedicamos ao desenvolvimento de métodos voltados para a análise da arquitetura genética de caracteres complexos, no mesmo estilo dos que seriam necessários quando tivéssemos acesso aos dados do nosso experimento. Para isso, nós utilizamos um banco de dados antigo, do laboratório do Prof. James Cheverud, que tinha uma estrutura parecida com os dados que viriam do nosso experimento. Com esses dados, exploramos novas técnicas de

⁹<https://f1000.com/prime/725409260>.

mapeamento genético desenhadas para aferir o padrão de efeitos genéticos em vários caracteres simultaneamente e relacioná-los com a história evolutiva da população estudada. Mais uma vez, meu conhecimento de estatística Bayesiana foi fundamental e eu desenhei toda a estratégia de análise, contando com a ajuda preciosa do Prof. Jason. Este trabalho resultou num manuscrito que se tornou um dos capítulos da minha tese, e foi publicado em Melo et al. (2019).

Além dessas visitas a Bath, enquanto estive no Brasil, distante do experimento com os camundongos, procurei me envolver com projetos em colaboração com os outros membros do LEM. Durante o ano de 2015, comecei um projeto com uma aluna de mestrado do LEM, Anna Penna, que hoje é doutora pela Universidade do Texas em San Antonio. Nesse período, Anna estava envolvida em um projeto grande de mestrado e tinha um conjunto de dados coletados durante a iniciação que não haviam sido explorados com cuidado. Eram dados de morfologia craniana de cinco populações experimentais de camundongos derivadas da mesma população ancestral, sendo duas populações selecionadas para aumento de peso, duas selecionadas para diminuição de peso, e uma controle. Trabalhando junto com a Anna, eu percebi que esse conjunto seria ideal para o estudo da evolução da covariação craniana sob seleção direcional, novamente um modelo experimental que se relacionava com minhas simulações feitas durante o mestrado. Num trabalho a quatro mãos, eu e Anna analisamos estes dados sob a perspectiva da covariação fenotípica, e encontramos um claro padrão de mudança na organização da variação das linhagens selecionadas, compatível com as previsões do meu trabalho de mestrado. Este trabalho também é um dos capítulos da minha tese e foi publicado em, Penna et al. (2017), sendo que eu e Anna aparecemos como co-primeiro autores.

Além da colaboração com a Anna, em 2015 eu também tive o privilégio, junto com o Prof. Gabriel, Prof. James Cheverud e um ex-aluno de mestrado do LEM, o Dr. Arthur Porto, atualmente professor na Universidade da Florida, de escrever uma revisão para um dos periódicos da *Annual Reviews*. O Gabriel havia sido convidado para escrever um artigo sobre modularidade e convidou nós três para participar. O Prof. Cheverud ficou responsável pela elaboração de uma introdução histórica ao conceito de modularidade, enquanto eu e Arthur lideramos a elaboração da maior parte do manuscrito e da produção de figuras, e o Gabriel contribuiu propondo ideias importantes ligadas às consequências macroevolutivas da teoria que nós revisamos no manuscrito. Eu, Arthur e Gabriel escrevemos a primeira versão do manuscrito em pouco mais de um mês de trabalho intenso, mas muito gratificante. Conseguimos apresentar um resumo didático da aplicação da teoria de modularidade em diversos níveis de organização biológica e, usando ideias do Gabriel, propor um papel maior para a modularidade e covariação genética na diversificação macroevolutiva. Este trabalho foi mais um dos capítulos da minha tese Melo et al. (2016).

Finalmente, em 2017, depois de uma série de atrasos devido às dificuldades técnicas de se criar uma colônia de camundongos tão grande, começamos a ter os primeiros resultados do trabalho com o Prof. Jason. Utilizei as ideias que desenvolvemos durante minha segunda visita a

Bath para analisar os dados disponíveis naquele momento, e produzi um conjunto de resultados preliminares que relacionava o padrão de efeitos genéticos nos fenótipos de curva de crescimento dos nossos camundongos com a história evolutiva das linhagens fundadoras. Usando um modelo de simulação mais complexo do que o que utilizei no mestrado, consegui também elucidar como o padrão de seleção nas linhagens fundadoras definia os efeitos genéticos que observamos na nossa geração F6. Este trabalho foi apresentado no Congresso Internacional de Evolução em 2018, e se tornou um dos capítulos da minha tese. Esse último capítulo não foi publicado, em parte porque ao mesmo tempo em que eu o escrevia, em setembro de 2018, nasceu meu primeiro filho, o Tomás, um mês antes do planejado!

Todas essas linhas de pesquisa e bancos de dados diferentes se juntaram numa tese coesa, que explorava de diferentes formas o problema da evolução das associações genéticas e suas consequências evolutivas. A necessidade de me envolver em diversos projetos independentes me obrigou a idealizar um fio condutor único nos trabalhos e, desenvolvendo técnicas adequadas a cada uma das perguntas abordadas nos trabalhos, consegui criar uma linha de pesquisa original e diversa. Enquanto o laboratório do Gabriel sempre trabalhou muito bem com a descrição e compreensão de padrões de covariação genética em mamíferos numa escala macroevolutiva, meu trabalho estendeu essa linha de pesquisa para um lado mais genético, trazendo novas técnicas e colocando a complexidade da arquitetura genética no centro da discussão. Processos evolutivos moldam a arquitetura genética, e a arquitetura genética afeta como as populações respondem a processos evolutivos, num processo de *feedback* complexo. Explorar essa interação se tornou o cerne da minha pesquisa.

Pós-doutorado

Laboratório de Evolução de Mamíferos

Quando estávamos discutindo meu projeto de doutorado, o Gabriel acreditava que eu deveria medir fenótipos cranianos com alta resolução nos camundongos da nossa população experimental. Os dados cranianos, aliados aos genótipos, permitem elucidar a arquitetura genética responsável pela variação craniana, e entender como um fenótipo complexo responde aos processos microevolutivos. Esse seria então um banco de dados rico, fornecendo substrato para uma série de projetos que poderiam estender de forma original as linhas de pesquisa do LEM. Entretanto, esse foi um ponto em que discordamos, não porque eu não tinha interesse nos dados cranianos, mas porque considerava que não teríamos tempo suficiente para completar as medições durante o doutorado. Assim, tomei a decisão de não incluir essa empreitada em nenhum dos projetos submetidos durante esse período, preferindo me concentrar nos fenótipos que poderiam ser coletados concomitantemente ao desenvolvimento da colônia: taxas de crescimento e composição corporal. Fiz isso confiante que esses dados seriam suficientes para o trabalho durante o doutorado e que obter os dados cranianos completos em tão pouco tempo seria muito difícil devido à escala do projeto. Porém, já planejando uma continuidade do projeto,

nos dois últimos anos de doutoramento comecei a digitalização os crânios dos animais da F6 no microtomógrafo de alta resolução do Departamento de Genética e Biologia Evolutiva.

Retrospectivamente, adiar o início de um projeto ligado aos dados cranianos foi uma escolha acertada. Escanear os animais foi um processo trabalhoso e cheio de imprevistos. Os animais chegaram ao Brasil no começo de 2017, após um processo demorado para garantir que eles passassem pela alfândega sem nenhum problema. Nessa época, eu coorientava uma aluna de iniciação fantástica, a Beatriz Demasi, que me ajudou muito no processo de importação dos animais do Reino Unido. Eu e a Beatriz passamos algumas semanas processando os cerca de 2400 animais conservados em álcool, catalogando e acondicionando-os em recipientes separados por geração do experimento. Depois disso, foram necessários alguns meses até aprendermos juntos a melhor forma de preparar e escanear os animais, respeitando as limitações de tamanho dos exemplares e tempo de uso do microtomógrafo do departamento. Tudo correu razoavelmente bem até que, em junho de 2017, após cerca de 250 animais escaneados, o microtomógrafo quebrou, e ficou quebrado durante mais de seis meses. Durante esse tempo, Beatriz, eu e Gabriel escrevemos um projeto de iniciação para ela, mas após duas tentativas não conseguimos financiamento FAPESP, por conta de problemas no currículo da aluna no início da graduação, que não refletiam em absoluto a qualidade da pesquisadora que ela com certeza vai se tornar. Depois disso, a Beatriz cansou de esperar o microtomógrafo e uma bolsa, e deixou o laboratório, mas não sem antes me ajudar a treinar minha outra aluna de iniciação, a Amanda Magalhães, no processo de preparação e escaneamento dos camundongos. No início de 2018, o microtomógrafo voltou finalmente a operar e, aproveitando a universidade vazia nas férias, consegui escanear mais de 1200 animais em 2 meses, praticamente sozinho. Depois dessa empreitada, com a demanda de uso do microtomógrafo aumentando, reduzimos o ritmo de digitalização dos animais, e ao longo de 2018 eu e Amanda terminamos de escanear todos os animais das gerações F5, F6 e alguns dos fundadores, num total de mais de 2200 animais escaneados¹⁰. Todo esse projeto, e principalmente o processo de obter os fenótipos cranianos, me ensinou a planejar no curto e no longo prazo, a escrever projetos resilientes, e a lidar com os imprevistos e atrasos que são comuns a praticamente todos os projetos dessa escala.

Coloco esse relato nesta seção porque o trabalho com os crânios se tornou o projeto do meu primeiro pós-doutorado, cujo objetivo é usar nossa colônia para investigar como a arquitetura genética se relaciona com o padrão de associações genéticas entre os traços do crânio. Como mencionei anteriormente, os dados cranianos permitem integrar este projeto com os outros resultados produzidos no LEM ao longo da última década, e explorar experimentalmente como as associações genéticas se estabelecem e evoluem num fenótipo complexo que teve sua história macroevolutiva amplamente estudada. Em 2020, três anos e meio após a chegada dos animais ao Brasil, finalmente concluímos a coleta dos fenótipos cranianos de todos os animais escaneados e

¹⁰ Até onde consegui averiguar, a maior amostra de morfologia craniana já feita em uma única população.

podemos começar a relacioná-los com os genótipos, a fim de identificar os loci responsáveis pela covariação no crânio.

O primeiro trabalho que usou o banco de dados dos crânios foi o projeto de iniciação da Amanda, que orientei durante o ano de 2019, quando somente a geração F6 havia sido medida. Amanda investigou a ligação entre a taxa de crescimento máxima observada nos animais da F6 com a sua integração¹¹ no crânio, duas variáveis que, quando medidas em diferentes espécies de mamíferos, aparecem positivamente correlacionadas numa escala macroevolutiva (Porto et al., 2013). O objetivo do projeto da Amanda foi verificar se essa associação era movida por uma ligação mecanística entre as duas variáveis, podendo, portanto, ser observada dentro de uma população com variação em taxas de crescimento. Nossa população era ideal para esse objetivo, pois fora fundada por linhagens com diferenças grandes nas suas taxas de crescimento estabelecidas por seleção artificial, e, portanto, apresentava muita variação em crescimento na geração F6. Esta pesquisa retomou uma das primeiras questões que eu e Gabriel nos colocamos ao começar o projeto conjunto com o Prof. Jason Wolf. Surpreendentemente, a relação entre taxa de crescimento e a integração era inversa ao observado na comparação entre espécies. Animais com taxas de crescimento mais altas se mostraram menos integrados que animais com a taxa de crescimento mais baixa. Apesar de não confirmar nossa hipótese inicial, esse resultado mostra uma ligação entre essas duas variáveis que ainda pretendemos explorar mais, com o objetivo de entender qual a consequência dessa associação para a evolução de curto prazo da covariação craniana. A orientação da Amanda envolveu um longo estudo de estatística, e ao final ela foi capaz de entender e interpretar o modelo Bayesiano de análise que construímos juntos. Apesar de ser um problema bem específico, as técnicas que utilizamos têm aplicação geral e podem ser usadas para analisar diversos tipos de dados experimentais.

No início de 2020, antes das medidas de contenção ocasionadas pela pandemia de covid-19, me dediquei a consolidar e fazer um controle de qualidade em todos os dados fenotípicos e genotípicos da colônia que construímos durante o meu projeto de doutorado. Uma das vantagens da coleta de fenótipos cranianos usando o microtomógrafo é a que, depois de medir manualmente todos os indivíduos usando as imagens digitalizadas, é possível voltar e encontrar erros e inconsistências nas medidas, que podem assim ser corrigidas. Por ser um processo muito trabalhoso, poucos grupos de pesquisa dedicam o tempo e cuidado necessários para produzir um banco de dados robusto e com o menor erro de medida possível. Mas eu acredito que, quando estamos tratando de medidas que podem ter uma fração de milímetro, esse cuidado é crucial para termos confiança nas análises. Para isso, eu desenvolvi um protocolo de verificação das medidas de crânio que permite detectar erros de medida muito pequenos, e usando esse protocolo produzi um conjunto de medidas muito precisas e acuradas da variação craniana desses animais. Esse protocolo foi aplicado a todos os animais escaneados, e os dados foram consolidados e organizados em um banco de dados que pode ser usado para uma série de projetos futuros.

¹¹Uma medida da intensidade média das associações entre um conjunto de traços fenotípicos.

Além disso, concomitantemente, o Prof. Arthur Porto estava trabalhando em técnicas de marcação automática das imagens de microtomógrafo utilizando métodos de aprendizado de máquina. Esses métodos permitem que, ao invés de medir manualmente os crânios, possamos usar um algoritmo para fazer as medidas automaticamente. Neste projeto, Arthur conseguiu desenvolver um algoritmo que mede automaticamente todos os pontos no crânio, e que pode ser usado para medir os crânios de todos os animais da colônia. A meu conjunto de dados, sendo de uma escala e qualidade inéditas, são ideais para a idealização e validação das técnicas de medidas automáticas que o Arthur desenvolve. Pensando nisso, estabelecemos uma colaboração potencialmente produtiva, que permitirá entender exatamente quais são as consequências e diferenças entre as medidas tradicionais e automáticas.

No momento sou o único pesquisador explorando os dados genéticos e fenotípicos da nossa colônia ativamente e, além das discussões com o Prof. Gabriel, mantenho contato com o Prof. Jason e os outros pesquisadores envolvidos no projeto inicial: Dra. Eleanor O'Brien, atualmente na Universidade de Bristol no Reino Unido, e o Prof. John Hunt, da Universidade de Exeter no Reino Unido. Esse banco de dados consolidado fornece um substrato rico para o estudo de arquitetura genética, covariação e a interação com processos microevolutivos, e portanto viabiliza uma série de projetos com perguntas específicas.

Alguns desses projetos já estão em andamento, como o de mapeamento da arquitetura genética e identificação de genes pleiotrópicos nos fenótipos de crescimento e cranianos, quantificação da relação entre arquitetura genética e covariação, ligação da arquitetura genética na população experimental com a história evolutiva dos fundadores, e a elucidação da base mecanística da ligação entre intensidade de integração e crescimento. Podemos utilizar nossos dados, aliados aos dados comparativos acumulados no LEM ao longo dos anos, para entender como se dá a ligação entre a micro e macroevolução de fenótipos morfológicos. Essa colônia permite também quantificar a contribuição de efeitos maternos para a variação em todos os fenótipos medidos, e entender as consequências evolutivas dessa variação via efeitos indiretos. Também podemos desenvolver métodos novos para o mapeamento genético de fenótipos de alta dimensão e explorar o padrão de herança da variação craniana. O tempo e esforço despendidos na produção desse banco de dados são certamente refletidos nas possibilidades de avanço do conhecimento que ele proporciona.

Além do meu projeto com os camundongos, durante meu período de pós-doutorado, continuei com nossas colaborações com outros membros egressos do laboratório. Destaco dois trabalhos desenvolvidos principalmente em colaboração com os Profs. Fabio Machado, da Oklahoma State University, e Prof. Alex Hubbe, da Universidade Federal da Bahia, além de outros colaboradores. Nestes trabalhos, nos debruçamos novamente sobre as consequências evolutivas da covariação em crânios de primatas, focando em diferenças metodológicas na medida da covariação entre caracteres morfológicos (Machado et al., 2019), e na influência do processo de desenvolvimento na manutenção dos padrões de covariação ao longo da vida dos indivíduos de diversos grupos

de mamíferos (Hubbe et al., 2022). Nesses trabalhos, novamente, participei desenvolvendo toda a parte estatística e metodológica, continuando o com o padrão de colaborações que estabeleci durante o doutorado.

A Pandemia e o Observatório Covid-19 BR

Em março de 2020, após finalizar a coleta de dados que descrevi acima, fui convidado pelo Prof. Paulo Inácio Prado a participar de um projeto de modelagem epidemiológica que ele estava coordenando, o Observatório Covid-19 BR¹². O objetivo do projeto era produzir projeções de curto prazo para a evolução da pandemia de covid-19 no Brasil, usando modelos epidemiológicos e dados de internações por síndrome respiratória aguda grave do SIVEP-GRIPE¹³ e posteriormente do eSUS notifica¹⁴. O projeto era uma colaboração entre o membros do Departamento de Ecologia do IB-USP, do Instituto de Física Teórica da UNESP, do Departamento de Matemática da Universidade Federal do ABC e de dezenas de outras universidades e institutos em vários países¹⁵. O projeto foi um sucesso, e as projeções do Observatório Covid-19 BR foram amplamente divulgadas na mídia e utilizadas por diversos órgãos públicos e privados para planejar ações de combate à pandemia. Minha participação foi centrada no desenvolvimento de métodos para a estimação de parâmetros epidemiológicos a partir dos dados de internações, e na análise dos dados de internações para entender a dinâmica da pandemia no Brasil. Sem algum tipo de tratamento estatístico, os dados de internação fornecem uma visão da pandemia que pode estar até 40 dias fora de compasso com a realidade. O principal produto do Observatório do qual que participei diretamente foi um conjunto de métodos para *nowcasting*, ou seja, entender o estado atual da pandemia corrigindo para o problema do atraso na atualização dos dados.

Ayroles Lab

Durante o doutorado e pós-doc na USP, estabeleci uma colaboração independente com o laboratório do Prof. Julien Ayroles na Universidade de Princeton nos Estados Unidos. O Prof. Ayroles lidera um grupo de pesquisa em genética, utilizando, entre outros modelos, uma colônia grande e diversa de *Drosophila melanogaster*, com foco em investigar interações epistáticas entre genes. O laboratório do Prof. Ayroles utiliza mapeamento genético, técnicas modernas de medição de fenótipos em larga escala, e outras técnicas de biologia de sistemas e genética quantitativa para entender a arquitetura genética por trás da variação em diversos tipos de fenótipos, tais como expressão gênica, morfologia, ou mesmo comportamentos. A colaboração com esse grupo me permite investigar aspectos da arquitetura genética que exigem amostras maiores do que seria viável com camundongos, como a influência da epistasia na covariação, interações entre fenótipo e ambiente, ou a covariação entre a expressão de vários genes simultaneamente. Como resultado dessa colaboração, no final de 2019, eu e o Prof. Ayroles

¹²<https://covid19br.github.io/>

¹³<http://info.gripe.fiocruz.br/>

¹⁴<https://datasus.saude.gov.br/notifica/>

¹⁵<https://covid19br.github.io/sobre>

decidimos submeter uma proposta para uma bolsa de pós-doutorado em Princeton, a *Princeton Presidential Postdoctoral Research Fellowship*. Assim, em julho de 2020 passei a integrar o grupo do Ayroles Lab. Minha contribuição para o grupo tem sido trazer uma visão não só da variação em cada traço independentemente, mas um foco em fenótipos multivariados e seu padrão de covariação. Essa abordagem multivariada permite estudar dinâmicas genéticas e evolutivas mais complexas, que revelam consequências que não podem ser acessadas investigando um traço de cada vez. Durante os cerca de 3 anos em Princeton, me evolvi em diversos projetos que detalho abaixo.

Contribuição da epistasia à resposta adaptativa

Enfrentei um desafio inicial nessa nova etapa de minha carreira: a impossibilidade de me mudar para Princeton devido às restrições na emissão de vistos americanos, uma consequência da pandemia. Isso me levou a participar de projetos em andamento no laboratório do Prof. Ayroles. Estabeleci minha primeira colaboração com o Dr. Simon Fosberg, um pós-doutorando focado em entender o impacto da epistasia - a interação entre dois ou mais loci genéticos - na evolução de características complexas. A epistasia, sendo um elemento crucial na arquitetura genética dessas características, é desafiadora de estudar devido à necessidade de grandes números amostrais para ser detectada com confiança.

Simon analisava dados de um experimento *Evolve & Resequence*¹⁶, envolvendo uma população *outbred*¹⁷ de *D. melanogaster*. A adaptação evolutiva, acreditamos, resulta da interação entre varreduras seletivas - variantes vantajosas que se tornam prevalentes rapidamente - e mudanças mais sutis em muitos loci, dependendo da arquitetura genética das características selecionadas.

No experimento, as moscas foram submetidas a dietas controle padrão e a dietas ricas em açúcar, como estresse ambiental, por mais de 100 gerações. Monitoramos as alterações na frequência alélica em mais de 4.000 moscas, identificando loci sob seleção. Os resultados apontaram que aproximadamente 4% do genoma estava sob seleção positiva, numa resposta majoritariamente poligênica. A maioria dos loci selecionados não indicava varreduras seletivas fortes, mas sim mudanças sutis na frequência alélica, alinhando-se a uma resposta poligênica. Identificamos também loci que evidenciavam epistasia, com mudanças na frequência alélica de um locus dependendo de outro, além de um padrão de desequilíbrio de ligação ao final do experimento.

Além de contribuir para a análise de dados, nesse projeto apliquei novamente a combinação de teoria, experimentos e simulações. Desenvolvi um modelo computacional que reproduzia nosso desenho experimental, avaliando a contribuição da epistasia para as mudanças nas frequências alélicas e no desequilíbrio de ligação. Comparando arquiteturas genéticas aditivas e epistáticas, demonstrei que apenas a epistasia explicava os padrões observados no experimento. Esses

¹⁶Em experimentos de E&R, uma população é exposta a uma pressão ambiental e tem suas frequências alélicas acompanhadas por sequenciamento a medida que se adapta ao novo ambiente.

¹⁷Geneticamente variável, em oposição às linhagens homozigotas normalmente utilizadas na pesquisa com *D. melanogaster*.

achados sugerem que a epistasia é um componente crucial na resposta evolutiva, e que as respostas poligênicas podem ser mais comuns do que se supunha anteriormente.

Apresentei esse trabalho no congresso da *Society for Molecular Biology & Evolution*, o SMBE2023, e na *Gordon Research Conference for Ecological and Evolutionary Genomics* 2023.

Mapeamento de eQTLs em *D. Melanogaster*

O próximo projeto que ataquei no Ayroles lab foi feito e colaboração com a Prof. Luisa Pallares, na época uma pós-doutorando e hoje professora do Max Plank de Biologia em Tübingen, na Alemanha. Nosso projeto visava trazer para *Drosophila* o tipo de catálogo de variantes que regulam expressão gênica disponível em outros organismos, como humanos e leveduras. *Drosophila* é um modelo para o qual a grande maioria do mapeamento genético é feito utilizando painéis de linhagens homozigotas, como o DGRP (Mackay et al., 2012) ou o DSPR (Long et al., 2014). Esses painéis são muito úteis para mapear efeitos aditivos, mas não permitem a detecção de efeitos de dominância ou epistáticos, fundamentais para entender os padrões complexos de interação entre loci diferentes. Para isso, é necessário uma população *outbred*, que é muito mais difícil de ser analisada, pois exige que todos os indivíduos num experimento sejam genotipados. O uso de uma população *outbred* geneticamente variável faz com que o poder estatístico não seja limitado pelo número de linhagens homozigotas disponíveis num dado painel. Ao contrário, podemos simplesmente amostrar quantos indivíduos forem necessários para um determinado experimento. Utilizando uma população de *Drosophila* geneticamente variável batizada de NEX, criada com o cruzamento de linhagens selvagens coletadas na Holanda, nós coletamos dados genéticos e de expressão gênica de mais de 1800 indivíduos, e mapeamos loci regulatórios de expressão (eQTLs) para mais de 5.000 genes. Com essa amostra inédita, foi possível mapear eQTLs em cis e trans para 98% dos genes expressos em *D. melanogaster*, aumentando em milhares o número de genes para os quais os loci regulatórios são conhecidos nesta espécie. Pela primeira vez, no contexto de uma população geneticamente variável, foram descritas as propriedades da regulação local e distal da expressão gênica em termos de diversidade genética, herdabilidade, conectividade e pleiotropia. Esses resultados elevam o entendimento da mosca da fruta a um nível que antes estava disponível apenas para poucos outros organismos, oferecendo um novo recurso de mapeamento que ampliará as possibilidades atualmente disponíveis para a comunidade de *Drosophila*. Estes recursos estão disponíveis em <https://drosophilaeqtl.org/> e o trabalho está disponível hoje como *pre-prints* (Pallares et al., 2023; Wolf et al., 2023) .

Modularidade em redes de co-expressão gênica

Um dos aspectos que mais me interessavam no trabalho com a Luisa era o padrão de co-expressão gênica, ou seja, a covariação observada entre expressão de genes. Trazendo minha experiência em covariação e sistemas multivariados, procuramos identificar os módulos de co-expressão nos dados que descrevi acima. Identificar esses módulos é um passo importante para reduzir a dimensionalidade da representação do transcriptoma, e permitir que a própria organização variacional do sistema informe os grupos de genes de interesse que podem ser analisados

separadamente, ou articulados para interpretar o padrão de regulação gênica. Uma dificuldade que encontramos nesse momento é de que mas as ferramentas disponíveis na literatura acabaram não lidando bem com o volume de dados do nosso experimento, resultando em agrupamentos pouco informativos e que excluíam a enorme maioria dos genes que analisamos. Procurando sanar esse problema, busquei inspiração na literatura de redes complexas, que nos levou a uma percepção importante: tradicionalmente, algoritmos de detecção de comunidades pressupõem que os genes estejam organizados em módulos assortativos, ou seja, que genes em um mesmo grupo estão mais associados entre si do que com genes em outros grupos. Contudo, essa abordagem pode ser arriscada, pois presume a existência desses módulos *a priori*, ignorando outras organizações possíveis das redes de co-expressão. Nos perguntamos, então, se seria possível encontrar comunidades biologicamente significativas sem impor uma organização modular às redes de co-expressão gênica, buscando esses blocos de genes por meio de uma medida de similaridade mais genérica. Para isso, adaptamos um método moderno de detecção de comunidades, os *Stochastic Block Models* (SBM), que não pressupõe a existência de módulos assortativos. Em vez disso, essa classe de modelos busca utilizar toda a informação contida na rede de coexpressão para separar os genes em blocos hierarquicamente organizados.

Utilizando dados de expressão gênica de RNA-seq medidos em dois tecidos derivados de uma população heterogênea de *Drosophila melanogaster*, os mostramos que: (a) o SBM consegue identificar dez vezes mais grupos do que métodos concorrentes, além de separar todos os genes em grupos (contra 10% dos métodos usuais); (b) vários desses grupos gênicos não são modulares, ou seja, os genes dentro de um grupo não são mais associados entre si do que com outros grupos, apesar de terem claro enriquecimento funcional para processos biológicos específicos; e (c) o enriquecimento funcional para grupos não modulares é tão forte quanto para grupos modulares. Esses achados indicam que o transcriptoma é estruturado de maneiras mais complexas do que se pensava anteriormente e sugerem a necessidade de revisar a suposição de que a modularidade é o principal fator na estruturação de redes de co-expressão gênica. Esse trabalho está disponível hoje como um *pre-print* (Melo et al., 2023).

Variabilidade da expressão gênica em humanos

Uma linha de pesquisa importante no Ayroles lab é o estudo de interações genótipo ambiente em humanos, especialmente no contexto de incompatibilidades genótipo ambiente, no qual populações humanas adaptadas a um ambiente são expostas a novas condições ambientais, seja por migração ou por mudanças no hábito de vida. Nesse contexto, uma pergunta fundamental é a variabilidade basal na expressão gênica em humanos. Existem genes com expressão mais ou menos variável, independente do ambiente? Ou a variabilidade em expressão gênica é completamente contexto dependente. Para abordar esse problema, em colaboração com um aluno de doutorado no laboratório, Scott Wolf, utilizamos conjuntos de dados de RNA-seq disponíveis publicamente para investigar a paisagem da variação na expressão gênica. Scott e eu levantamos mais de 50 estudos, abrangendo mais de 20 mil medidas de expressão gênica. Esses dados

abrangem uma variedade de tecidos e permitem avaliar se existem genes consistentemente mais ou menos variáveis entre os tecidos e conjuntos de dados, bem como os mecanismos que impulsionam esses padrões.

Os resultados mostram que a variação na expressão gênica é relativamente semelhante entre diferentes tecidos e estudos, indicando que o padrão de variação da transcrição é consistente. A partir desta semelhança, criamos rankings globais e específicos para cada tecido de variação, demonstrando que a função do gene, variação de sequência e assinaturas regulatórias de genes contribuem para a variação na expressão gênica. A criação desses rankings globais só foi possível devido à minha experiência em lidar com dados multivariados, pois nós nos valemos das técnicas matemáticas usada na genética quantitativa para desenvolver nosso método de ranqueamento global. Usando esse ranking, mostramos que os genes de baixa variação estão associados a processos celulares fundamentais, possuem menores níveis de polimorfismos genéticos, maior conectividade gene-gene, e tendem a estar relacionados a estados de cromatina associados à transcrição. Enquanto isso, genes de alta variação são enriquecidos em genes envolvidos na resposta imune e genes responsivos ao ambiente, e estão associados a níveis mais elevados de polimorfismos.

Esses resultados revelam que o padrão de variação da transcrição não é um ruído aleatório, mas sim uma característica gênica consistente que parece ser funcionalmente restrita em populações humanas. Além disso, esse aspecto comumente negligenciado da variação fenotípica molecular contém informações importantes para entender traços complexos e doenças. Esse trabalho foi publicado em Wolf et al. (2023), artigo em que apareço como co-primeiro autor.

Projetos em andamento e próximos passos

Na biologia contemporânea, a geração de conjuntos de dados deixou de ser o principal desafio. Com o avanço das tecnologias de sequenciamento de DNA e RNA, e de outras técnicas de alta precisão, a quantidade de dados genômicos e de fenótipos moleculares disponíveis cresceu exponencialmente. Este acúmulo de informações, embora seja um marco importante, trouxe uma complexidade sem precedentes na análise e compreensão dos dados. A única forma de lidar com essa complexidade é por meio de uma abordagem integrativa, que combine teoria, experimentos e análises computacionais. Acredito que minha formação e experiência me permitem contribuir com essa abordagem integrativa, e que o Instituto de Biociências da USP é o lugar ideal para desenvolver essa linha de pesquisa. Uma percepção importante é a quantidade de *insight* biológico acumulado nos anos de pesquisa em evolução morfológica e na ampla teoria de genética quantitativa desenvolvida no século XX. De fato, uma espécie de redescobrimiento da genética quantitativa foi um dos artigos mais influentes dos últimos anos (Boyle et al., 2017), destacando como muitas variantes de pequenos efeitos podem interagir para produzir fenótipos complexos. A genética quantitativa evolutiva é uma ferramenta poderosa para entender e esmiuçar a arquitetura genética de fenótipos complexos.

Nos últimos anos do meu pós-doutorado, meu foco foi em estabelecer uma linha de pesquisa independente, focada em levar nosso entendimento de redes de co-expressão gênica para um novo patamar. Para isso, estabeleci colaborações com grupos de pesquisa experimentais de ponta, como o laboratório do Prof. Ayroles em Princeton, e o laboratório do Prof. Pallares em Tübingen, entre outros que já descrevi aqui. Essas colaborações internacionais me permitiram desenvolver projetos em uma escala que não seriam possíveis no contexto brasileiro. Atualmente, estamos construindo um conjunto de dados que usa experimentos de seleção natural no laboratório, expondo populações a um estress ambiental por muitas gerações, com experimentos de exposição aguda, onde uma população é exposta ao estress por apenas uma geração. Esse tipo de experimento nos permite entender como a resposta de curto prazo, plástica, e a resposta de longo prazo, adaptativa, estão relacionadas. O meu interesse nesses experimentos é entender como a co-expressão gênica se relaciona com a resposta adaptativa, e como a epistasia influencia essa relação. Nós já mostramos como a co-expressão pode ter estruturas diferentes dependendo do contexto (Melo et al., 2023), e agora queremos entender como essa alteração se relaciona com a resposta adaptativa e plástica, e com a arquitetura genética da regulação gênica. O desenvolvimento de novos métodos para a análise de co-expressão certamente será fundamental para que o conhecimento gerado nesses projetos leve a um entendimento mais profundo dos mecanismos de regulação e adaptação relacionados à expressão gênica.

Continuo também com meu trabalho relacionado a arquitetura genética de caracteres morfológicos, que tem consequências importantes para o melhoramento animal e para a compreensão do processo macro-evolutivo. Para isso, utilizo os dados que coletei durante meu doutorado, e que descrevi acima, e continuo com colaborações com outros grupos interessados em problemas macro-evolutivos, como o laboratório do Prof. Fabio Machado, curador da coleção de mamíferos na Universidade Estadual de Oklahoma, e o Prof. Gabriel Marroig, do Departamento de Genética e Biologia Evolutiva do IB.

A construção desses conjuntos de dados e as colaborações já estabelecidas, além da experiência computacional, tanto com simulações quanto análises customizadas, me permitem continuar desenvolvendo uma linha de pesquisa de ponta e original, que procura entender a relação entre arquitetura genética, covariação e evolução multivariada. Essa linha de pesquisa se encaixa bem na estrutura geral do Instituto de Biociências, contribuindo com uma visão teórica que liga os processos genéticos com os processos evolutivos e que pode ser instrumental para as outras linhas de pesquisa. Além disso, os métodos estatísticos que uso e desenvolvo são relevantes para a genética humana, especialmente as técnicas modernas de mapeamento genético aplicadas à elucidação da arquitetura genética de caracteres complexos. Essas contribuições teóricas e metodológicas poderiam se dar não apenas por meio de colaborações com os outros docentes do instituto, mas também oferecendo disciplinas que irão treinar os alunos em teoria evolutiva ou nos métodos estatísticos, como detalho a seguir.

Ensino

Durante minha trajetória acadêmica, sempre procurei dedicar tempo ao ensino, seja contribuindo como monitor em disciplinas de graduação e pós-graduação ou oferecendo disciplinas, cursos e *workshops*. Como já mencionei, as monitorias foram muito importantes na minha formação, me permitindo ter contato com áreas e métodos diferentes daqueles com que eu trabalhava no laboratório. Acredito que essas experiências foram cruciais para me proporcionar uma visão holística e integrativa de perguntas biológicas, expandindo minha visão para além das escalas do organismo e das populações que me são mais familiares. Participei como monitor de várias disciplinas do Departamento de Ecologia, como a de programação em R, a de modelos estatísticos e, talvez a experiência mais formadora, o Curso de Campo da Mata Atlântica, organizado pelos Profs. Glauco Machado, Paulo Inácio Prado e Adriana Martini.

No curso de campo, levamos 20 alunos de mestrado para uma reserva ecológica em Barra do Una, ao sul de Peruíbe. Lá, sem acesso a internet ou sinal de celular, os alunos desenvolvem uma série de pequenos projetos de pesquisa e produzem manuscritos, que são revisados pelos professores e monitores. Ao longo de um mês, cada aluno participa de quatro projetos em grupo, que são propostos pela equipe da disciplina. Cada aluno também idealiza e executa um projeto individual, que é aprovado pela equipe de professores. Cada projeto é orientado por um professor convidado ou um monitor. É um curso muito intenso de desenho experimental, coleta e análise de dados, escrita e apresentação científica. Na edição em que eu participei, em 2015, apenas o professor Glauco, eu e um outro monitor passamos os 30 dias acompanhando os alunos, enquanto os outros professores e convidados passavam tipicamente uma semana de cada vez. Para mim, foi como uma experiência de orientação condensada, pois ajudei a produzir um total de 40 pequenos manuscritos, com até 5 revisões cada um. Aprendi muito sobre como revisar um texto, como orientar projetos num ambiente de alta pressão e como incentivar os alunos a fazerem o melhor trabalho possível. Foi um processo difícil para mim, no qual aprendi muito, tanto acertando, quanto errando. O momento mais desafiador do curso foi ajudar os alunos nos seus projetos individuais, já que eram 20 alunos produzindo projetos independentes, todos procurando orientação e ajuda com a teoria, análise e interpretação dos seus resultados. A demanda cognitiva de pensar e interpretar 20 projetos simultaneamente não é pequena, mas este foi um momento formativo para mim.

Contribuí também com as disciplinas da Genética, atuando como monitor da disciplina de Biologia Evolutiva ao longo de vários anos, tendo escrito junto com a Dra. Monique Simon uma apostila para a disciplina¹⁸. Mais recentemente eu também produzi todo o material de aula prática em R¹⁹ para a Biologia Evolutiva. Também fui monitor PAE da disciplina de Processos Evolutivos duas vezes ao longo da minha pós-graduação.

¹⁸<https://github.com/lem-usp/apostila-bio-evol/blob/master/apostila-Bio312.pdf?raw=true>.

¹⁹<https://diogro.github.io/BioEvol/>.

Durante o pós-doutorado na USP, ofereci uma disciplina de pós-graduação com a Dra. Monique Simon no Programa de Pós-Graduação em Biologia Comparada da FFCLRP-USP, em Ribeirão Preto, na qual nós apresentamos uma introdução à modularidade e teoria evolutiva e ligamos esses formalismos com suas consequências práticas. A disciplina conta com aulas teóricas e práticas e já foi oferecida duas vezes, em 2017 e 2019. Também estive envolvido, junto com vários pós-doutorandos de todos os departamentos do IB, numa disciplina prática de escrita científica. Apesar de ser uma das atividades mais frequentes e importantes da carreira científica, temos muito pouco treino objetivo de escrita durante a pós-graduação. Isso é uma pena, já que a escrita é a principal ferramenta de disseminação do conhecimento produzido em nossos laboratórios. Percebendo isso, eu dediquei um tempo considerável de estudo a melhorar minha escrita em inglês, focando principalmente na escrita de artigos e projetos. Nossa disciplina apresenta aos alunos maneiras de pensar a escrita científica e oferece métodos para que eles possam elaborar textos claros e cativantes.

Já em Princeton, participei como monitor de uma disciplina de Genética Humana, Bioinformática, e Biologia Evolutiva, ministrada pelos professores Mona Singh, Sarah Kocher e Julien Ayroles; e de uma disciplina de Epidemiologia, ministrada pelo Dr. Benjamin Rice, para a qual minha experiência no Observatório Covid-19 BR foi fundamental. No último ano, fui também responsável por desenvolver e ministrar uma disciplina de Computação para Biologia, que apresentava aos alunos de graduação de diversos departamentos os principais conceitos de programação e análise de dados em R e python, com um foco na pesquisa em ecologia e genômica.

Além do ensino em Princeton, mantenho meu esforço em formar alunos no Brasil participando do Programa de Formação em Ecologia Quantitativa²⁰, uma iniciativa do Instituto Serrapilheira que visa capacitar alunos de mestrado em métodos quantitativos modernos. Participei como instrutor de um curso de Modelagem Estatística, organizado em 2023 pelo Prof. Paulo Inácio Prado. Nesse curso, apresentei aos alunos os principais conceitos de modelagem estatística, com um foco em modelos lineares e generalizados, e mostrei como implementar esses modelos na prática. A partir de 2024, eu assumo a organização desse curso de modelagem estatística, em parceria com a Dr. Sara Mortara, da *Re.Green*²¹, uma ONG focada em conservação.

Com base nessas experiências didáticas, entendo que um bom curso de graduação ou de pós-graduação deveria oferecer as ferramentas para que os alunos possam se tornar capazes de idealizar, analisar e interpretar seus dados e experimentos de forma independente. Para isso, é fundamental oferecer tanto disciplinas instrumentais quanto teóricas. Nesse contexto, apresento algumas propostas de contribuições que eu poderia fornecer ao departamento, buscando proporcionar aos alunos do IB e da USP a melhor formação possível, voltada para o pensamento crítico, independência e rigor científico.

²⁰<https://serrapilheira.org/programa-de-formacao-em-ecologia-quantitativa/>

²¹<https://re.green/>

Teoria evolutiva

Das disciplinas de graduação atualmente oferecidas pelo departamento, as que eu me sentiria mais confortável para ministrar seriam aquelas que envolvem ensino de evolução, como Processos Evolutivos, Biologia Evolutiva, Abordagens Multidisciplinares em Genética, ou uma das disciplinas de Genética e Evolução. Poderia também contribuir com a disciplina de Introdução à Programação de Computadores para Biologia ou a Introdução à Bioinformática.

Como disciplina eletiva, eu ofereceria uma disciplina de Teoria Evolutiva, apresentando de forma mais profunda os fundamentos teóricos da biologia evolutiva, abordando não só as formulações tradicionais em termos de frequências alélicas de Fisher e Wright, mas também as formulações modernas como o Teorema de Price ou a genética quantitativa multivariada e as aplicações destes formalismos, como melhoramento animal, mapeamento genético e estudos evolutivos em populações naturais. Numa versão avançada para a pós-graduação, uma disciplina de teoria evolutiva poderia abordar também seleção multinível, modelos de manutenção da variação genética e eventualmente, para alunos com algum conhecimento de cálculo, até mesmo a formulação usando equações diferenciais parciais da evolução de frequências alélicas iniciada por Kimura.

Estatística na era da computação

Outra contribuição que eu poderia oferecer na graduação ou na pós-graduação seria uma disciplina de análise de dados utilizando ferramentas computacionais modernas e modelos Bayesianos. Na pós-graduação, essa disciplina poderia ser planejada em conjunto com as disciplinas de programação em R e modelagem estatística do Departamento de Ecologia, fornecendo aos alunos de pós do IB um caminho completo para uma formação moderna e adaptável em métodos estatísticos que os capacita para realizarem suas pesquisas com maior autonomia. Uma vantagem de oferecer esse tipo de disciplina no Instituto de Biociências é poder enfatizar aplicações em genética, como mapeamento genético ou modelos evolutivos e demográficos que não são incluídos nos cursos da matemática. Esses métodos estatísticos computacionais e de bioinformática são cada vez mais relevantes no estudo da genética, com a disseminação de dados genômicos de larga escala que exigem análises sofisticadas e específicas.

Atividades administrativas

Desde o mestrado, tive uma participação constante como representante discente (RD) em diversas comissões dentro do IB. Minha entrada nesse aspecto da vida na universidade não se deu por uma vocação para participar de atividades administrativas nem por um interesse pela política universitária, mas pela percepção pragmática de que alguém tinha que fazer esse papel. Uma colega da graduação envolvida com o centro acadêmico, a Flávia Ferrari, me convidou para ser candidato a RD na Congregação do IB, pois ela achava que eu poderia ser um bom representante mas, principalmente, por que não haviam outros candidatos interessados. Então, sabendo que era eu ou ninguém, aceitei o convite.

Participar da Congregação acabou sendo muito informativo, e pude ver de perto como a universidade funcionava e como os professores navegavam a política universitária. Foi nesse momento que eu percebi que a administração universitária tem consequências concretas para os alunos, professores e para as atividades fim da universidade. No meu primeiro mandato como RD na Congregação, pude acompanhar de perto a importância da representação discente como mecanismo de interlocução entre estudantes, docentes e as instâncias administrativas. Por exemplo, me lembro de um caso em que dois alunos tinham demandas essencialmente idênticas (uma extensão de prazo de algum tipo), mas um dos alunos contatou os RDs e explicou o motivo de seu pedido, enquanto o outro apenas enviou os requerimentos necessários. O primeiro teve a sua demanda atendida, o segundo não. A percepção da importância dessa atividade fez com que, depois dessa primeira experiência, eu continuasse atuando como RD em diversas comissões ao longo de toda a pós-graduação. No mestrado, além da Congregação participei também como RD na Comissão de Pós-Graduação. Já no doutorado, participei novamente da Congregação e da Comissão de Pós-Graduação, depois fui eleito para a Comissão de Pesquisa por dois mandatos, e finalmente, representei os alunos da genética na Comissão Coordenadora de Programa (CCP) do Programa de Pós-Graduação em Ciências Biológicas (Biologia Genética) por 2 mandatos, ficando como representante até o último mês do meu doutorado. Durante essa trajetória nas comissões, começando na Congregação e descendo a hierarquia de comissões que regem a pós-graduação até a CCP, procurei adotar uma participação construtiva e propositiva que contribuísse para a formação dos alunos, atuando como uma ponte entre seus interesses e necessidades e a realidade do instituto.

Referências

- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 7, 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- Costa, T. M., Hebets, E. A., Melo, D., & Willemart, R. H. (2016). Costly learning: preference for familiar food persists despite negative impact on survival. *Biology Letters*, 7, 20160256. <https://doi.org/10.1098/rsbl.2016.0256>
- Fukimoto, N., Melo, D., Palme, R., Zanella, A. J., & Mendonça-Furtado, O. (2020). Are cats less stressed in homes than in shelters? A study of personality and faecal cortisol metabolites. *Applied Animal Behaviour Science*, 104919. <https://doi.org/10.1016/j.applanim.2019.104919>
- Hubbe, A., Machado, F. A., Melo, D., Garcia, G., Sebastião, H., Porto, A., Cheverud, J., & Marroig, G. (2022). Morphological integration during postnatal ontogeny: implications for evolutionary biology. *Evolution*. <https://doi.org/10.1093/evolut/qpac052>
- Hubbe, A., Melo, D., & Marroig, G. (2016). A case study of extant and extinct Xenarthra cranium covariance structure: implications and applications to paleontology. *Paleobiology*, 3, 465–488. <https://doi.org/10.1017/pab.2015.49>
- Long, A. D., Macdonald, S. J., & King, E. G. (2014). Dissecting complex traits using the *Drosophila* Synthetic Population Resource. *Trends Genet.*, 11, 488–495. <https://doi.org/10.1016/j.tig.2014.07.009>

- Machado, F. A., Hubbe, A., Melo, D., Porto, A., & Marroig, G. (2019). Measuring the magnitude of morphological integration: The effect of differences in morphometric representations and the inclusion of size. *Evolution*, 12, 2518–2528. <https://doi.org/10.1111/evo.13864>
- Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., Richardson, M. F., Anholt, R. R. H., Barrón, M., Bess, C., Blankenburg, K. P., Carbone, M. A., Castellano, D., Chaboub, L., Duncan, L., ... Gibbs, R. A. (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, 7384, 173–178. <https://doi.org/10.1038/nature10811>
- Marroig, G., Melo, D., & Garcia, G. (2012). Modularity, noise, and natural selection. *Evolution*, 5, 1506–1524. <https://doi.org/10.1111/j.1558-5646.2011.01555.x>
- Melo, D., & Marroig, G. (2015). Directional selection can drive the evolution of modularity in complex traits. *Proceedings of the National Academy of Sciences*, 2, 470–475. <https://doi.org/10.1073/pnas.1322632112>
- Melo, D., Garcia, G., Hubbe, A., Assis, A. P., & Marroig, G. (2015). EvolQG - An R package for evolutionary quantitative genetics. *FI000Research*, 925. <https://doi.org/10.12688/fi000research.7082.3>
- Melo, D., Marroig, G., & Wolf, J. B. (2019). Genomic Perspective on Multivariate Variation, Pleiotropy, and Evolution. *Journal of Heredity*, 4, 479–493. <https://doi.org/10.1093/jhered/esz011>
- Melo, D., Pallares, L. F., & Ayroles, J. F. (2023). Reassessing the modularity of gene co-expression networks using the Stochastic Block Model. *bioRxiv*, 2023. <https://doi.org/10.1101/2023.05.31.542906>
- Melo, D., Porto, A., Cheverud, J. M., & Marroig, G. (2016). Modularity: genes, development and evolution. *Annual Reviews of Ecology, Evolution, and Systematics*, 1, 463–486. <https://doi.org/10.1146/annurev-ecolsys-121415-032409>
- Pallares, L. F., Melo, D., Wolf, S., Cofer, E. M., Varada, V., Peng, J., & Ayroles, J. F. (2023). Saturating the eQTL map in *Drosophila melanogaster*: genome-wide patterns of cis and trans regulation of transcriptional variation in outbred populations. *bioRxiv*, 2023. <https://doi.org/10.1101/2023.05.20.541576>
- Penna, A., Melo, D., Bernardi, S., Oyarzabal, M. I., & Marroig, G. (2017). The evolution of phenotypic integration: How directional selection reshapes covariation in mice. *Evolution*, 10, 2370–2380. <https://doi.org/10.1111/evo.13304>
- Porto, A., Shirai, L. T., Oliveira, F. B. de, & Marroig, G. (2013). Size variation, growth strategies, and the evolution of modularity in the mammalian skull. *Evolution*, 11, 3305–3322. <https://doi.org/10.1111/evo.12177>
- Wolf, J. B., Howie, J. A., Parkinson, K., Gruenheit, N., Melo, D., Rozen, D., & Thompson, C. R. L. (2015). Fitness Trade-offs Result in the Illusion of Social Success. *Current Biology*, 8, 1086–1090. <https://doi.org/10.1016/j.cub.2015.02.061>
- Wolf, S., Abhyankar, V., Melo, D., Ayroles, J. F., & Pallares, L. F. (2023). From GWAS to signal validation: An approach for estimating genetic effects while preserving genomic context. *bioRxiv*, 2023. <https://doi.org/10.1101/2023.03.09.531909>

Wolf, S., Melo, D., Garske, K. M., Pallares, L. F., Lea, A. J., & Ayroles, J. F. (2023). Characterizing the landscape of gene expression variance in humans. *PLoS Genetics*, 7. <https://doi.org/10.1371/journal.pgen.1010833>