

Willian Domingues Coelho

**Soluções computacionais para a coleta e análise de dados em programas de
melhoramento genético animal**

Dissertação apresentada ao Programa de Pós-Graduação em Zootecnia da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Zootecnia.

Orientadora: Dra. Arione Augusti Boligon

Co-orientador: Dr. Fernando Flores Cardoso

Pelotas, 2017
Willian Domingues Coelho

Ferramentas para otimização da coleta e análise de dados em programas de melhoramento genético animal

Dissertação aprovada, como requisito parcial, para obtenção do grau de Mestre em Zootecnia, Programa Pós-Graduação em Zootecnia, Universidade Federal de Pelotas.

Data da defesa: 16 de fevereiro de 2017.

Banca examinadora:

Prof. Dr. Arione Augusti Boligon (Orientadora)
Doutora em Genética e Melhoramento Animal pela Universidade Estadual Paulista
Júlio de Mesquita Filho - UNESP

Prof. Dr. Marcos Jun-Iti Yokoo
Doutor em Genética e Melhoramento Animal pela Universidade Estadual Paulista
Júlio de Mesquita Filho – UNESP

Prof. Dr. Vinícius do Nascimento Lampert
Doutor em Zootecnia UFRGS - Universidade Federal do Rio Grande do Sul

Prof. Dr. Leandro Lunardini Cardoso
Doutor em Zootecnia pela UFRGS - Universidade Federal do Rio Grande do Sul

Resumo

Diferentes programas de melhoramento genético animal estão surgindo e se consolidando ao redor do mundo. Com isso, técnicas para análise de dados coletados são constantemente aprimoradas para aumentar a eficácia dos resultados. Contudo, com os avanços da tecnologia na área computacional e com a popularização dos dispositivos móveis (*smartphones* e *tablets*) e plataformas web, surgiram diversas oportunidades de modernizar e automatizar ainda mais o processo de avaliação e seleção animal. Este trabalho apresenta o desenvolvimento de dois *softwares* que visam aprimorar a forma como dados são analisados e, para um programa de melhoramento animal específico, como são coletados. No primeiro, é apresentado um aplicativo para dispositivos móveis que facilita a busca e o envio de dados sobre bovinos avaliados geneticamente no programa de melhoramento animal PampaPlus. Esse programa é realizado em parceria entre a Associação Brasileira de Hereford e Braford e a Embrapa Pecuária Sul na região Sul do Brasil. O segundo trata-se da integração entre uma ferramenta em linguagem R, que calcula o Desequilíbrio de Ligação entre populações (*LD Estimate*) para o ambiente do *Galaxy Project* com intuito de simplificar e popularizar o seu uso.

Palavras-Chave: análise de dados, coleta e acesso de dados, desenvolvimento de *softwares*; desequilíbrio de ligação; Galaxy Project

Abstract

Different animal breeding programs are emerging and consolidating around the world. Thus, techniques for analyzing collected data are constantly improved to increase the effectiveness of the results. However, with advances in computer technology and the popularization of mobile devices (tablets and cell phones), emerged several opportunities to modernize and automate the process of animal selection. This work presents the development of two softwares that aim to improve the way data are analyzed and, for a specific animal breeding program, how they are collected. In the first, a mobile application is presented that facilitates the search and submission of data on genetically evaluated cattle in the PampaPlus animal breeding program. This program is carried out in partnership between the Brazilian Association of Hereford and Braford and Embrapa Pecuária Sul in the southern region of Brazil. The second one is the integration between an R-language tool, which calculates the linkage disequilibrium between populations (LD Estimates) for the Galaxy Project environment in order to simplify and popularize its use.

Key-Words: data analysis, data collection and access, Galaxy Project; linkage disequilibrium; software development

Sumário

1 Introdução Geral	9
2 Objetivos	13
2.1 Objetivos gerais	13
2.2 Objetivos específicos	13
3 PampaPlusApp: Ferramenta para otimizar a coleta de dados utilizados em avaliações genéticas	14
3.1 Introdução	14
3.2 Escolha da plataforma	15
3.3 Desenvolvimento	16
3.4 Principais Funcionalidades	33
3.5 Estágio atual	33
3.6 Downloads e testes em dispositivos reais	34
3.6.1 Armazenamento	34
3.7 Comparação entre coleta de dados via PampaPlusApp e convencional	37
3.8 Trabalhos Futuros	39
4 Easy LD-Estimates: adaptação do software <i>LD-Estimates</i> a plataforma <i>Galaxy Project</i>	40
4.1 Introdução	40
4.2 Desenvolvimento	42
4.3 Execução de análise para validação da adaptação	49
4.4 Benefícios observados	58
5 Considerações Finais	60
Referências	61

1 Introdução Geral

Sabe-se que o emprego de técnicas que promovem o melhoramento genético de bovinos tem crescido com o passar dos anos, resultando em maior rapidez e eficiência na obtenção de ganhos genéticos. Sendo assim, para que essas técnicas sejam aplicadas corretamente, é necessário o maior nível de precisão possível, tanto no que diz respeito aos dados utilizados, quanto à análise dos mesmos. Contudo, ainda que os programas de avaliação genética estejam em constante aprimoramento, diversos problemas dificultam partes cruciais do processo, acarretando em dispêndio significativo de tempo e diminuição da acurácia dos resultados. Com relação a redução da acurácia dos resultados, um dos principais fatores agravantes está na coleta e transmissão de dados.

A coleta de dados é uma das principais etapas dos programas de melhoramento, visto que erros nesse processo comprometem todos os passos subsequentes, bem como o sucesso de todo o programa. Alguns destes problemas estão diretamente relacionados com os métodos empregados na coleta. Esses métodos, muitas vezes ineficientes e totalmente manuais, seja por falta de tecnologia de apoio ao produtor ou desconhecimento da mesma, implicam em uma queda na confiança dos dados coletados. Outros fatores, como a baixa adaptação dos técnicos e/ou produtores rurais às ferramentas computacionais existentes, as limitações das propriedades rurais em relação à conexão com a internet e, inclusive, a ocasional falta de energia elétrica, são os principais fatores responsáveis por uma diminuição na acurácia dos dados. Além disso, existe certa resistência na utilização de novas tecnologias por parte de alguns produtores, desta forma, as novas ferramentas devem ser simples, rápidas, atrativas e com interface amigável para que sejam aceitas mais rapidamente no mercado. Um exemplo de programa de melhoramento animal que possui as características citadas é o PampaPlus.

O programa de avaliação de bovinos de corte PampaPlus, desenvolvido pelo Laboratório de Bioinformática e Estatística Genômica (Labegen) da Embrapa Pecuária Sul em parceria com a Associação Brasileira de Hereford e Braford (ABHB), está consolidado. Contudo, existem certos fatores que podem aumentar ainda mais a acurácia dos resultados e tornar o programa ainda mais confiável. Atualmente, a principal preocupação em relação a esse programa é o quão correto estão os dados que são informados pelos produtores participantes do mesmo, pois

sabe-se que em diversas propriedades rurais ainda não existe conexão com a internet e, nas que existe, ocorre com velocidade reduzida ou com falhas constantes. Além disso, como os dados do programa hoje são informados apenas pela ferramenta online PampaPlusNet (www.pampaplusnet.com.br), entre a coleta e o envio dos mesmos existem diversos fatores que dificultam o uso eficaz do sistema.

No programa PampaPlus a coleta de dados acaba sendo constituída por um maior número de etapas do que deveria, sendo uma coleta no centro de manejo e posterior alimentação do sistema, com base nas anotações realizadas. Dessa forma, as informações são obtidas de modo não otimizado, possibilitando a entrada de erros. Durante a coleta dos dados podem ocorrer, com frequência, erros de digitação e de leitura de equipamentos. Ainda nos casos em que são feitas anotações manuais antes da inserção no sistema, podem ocorrer erros de leitura do que foi escrito no papel no qual são realizadas as anotações, pois geralmente a pessoa que coleta os dados no campo não é a mesma que os insere no sistema. Logo, uma forma de melhorar o desempenho da coleta de dados seria disponibilizar uma ferramenta que trabalhe *off-line*, de modo que o produtor rural possa registrar os dados diretamente de forma digital, simples, atrativa, e sem que a dependência de uma fonte de alimentação elétrica. Isso facilitaria a adaptação dos produtores e aumentaria a aceitação dos mesmos a essa nova tecnologia.

Sabendo-se dessa necessidade dos produtores de uma ferramenta *off-line* e, em virtude do crescente aumento do número de dispositivos móveis com conexão a internet (como *tablets*, celulares, entre outros), e consequente popularização deste tipo de equipamento, uma possível solução para facilitar a coleta de informações seria a produção de um aplicativo para dispositivos móveis. Esses dispositivos móveis incluem *tablets* e *smartphones*, que podem possuir alguns recursos quando conectados a internet, além de outras funcionalidades independentemente do acesso à internet. Com isso, este aplicativo poderia ser utilizado no momento da coleta de dados, que geralmente ocorre em ambientes em que o uso de *laptops* seria inviável, seja por limitações de espaço, indisponibilidade de conexão com a internet ou fonte de energia elétrica. Este aplicativo poderia armazenar as novas informações e fazer o upload para o servidor principal automaticamente quando uma rede estivesse disponível. Esse tipo de desenvolvimento de *software* se

justifica devido ao imenso crescimento na indústria de dispositivos móveis nos últimos anos, mostrando que uma nova era de comunicação se instalou.

No Brasil, existem aproximadamente 283,52 milhões de linhas telefônicas registradas, sendo 6,3 milhões de dispositivos com acesso a internet 4G, o que caracteriza dispositivos de última geração, apesar de poucas cidades do país contarem com esta tecnologia (ANATEL). Isso gera uma constante procura por aplicativos para *smartphones* ou *tablets*, o que nos mostra que os usuários, em sua imensa maioria, já não querem ficar presos a aplicações que funcionem somente em computadores *desktop* e *laptops*. Essa tendência se confirma quando analisamos os dados de vendas, enquanto a venda de *laptops* está em queda, a de *smartphones* não para de quebrar recordes. O principal fator que causou essa migração de computadores para *tablets* e celulares foi a mobilidade. As conexões de rápido acesso com internet 3G, 4G e redes Wi-Fi permitem que os dispositivos fiquem sempre conectados e prontos para serem empregados nas mais diversas funções. Dentre os principais sistemas operacionais disponíveis no mercado, três deles possuem 99% dos usuários do planeta, sendo 82% usuários de Android (Google), 15% de iOS (Apple) e cerca de 2,9% de Windows Phone (Microsoft).

Outro benefício da construção de aplicativos para dispositivos móveis é proporcionar um novo modo de acesso aos dados pelo usuário. O sistema atual do PampaPlusNet permite que o usuário possa realizar o acesso através de seu *tablet* ou celular, contudo, é necessário ter uma conexão com a internet. Além disso, como o sistema foi desenvolvido e otimizado com foco na plataforma PC, a alimentação de dados via celular não é a ideal. Desta forma, o aplicativo não só otimizaria o modo como os dados são coletados, como também permitiria que as informações fossem acessadas mesmo quando o dispositivo não estivesse conectado à internet. Possibilitando, assim, uma economia no pacote de internet do usuário e prevenindo uma possível cobrança extra por parte da operadora, caso o limite seja ultrapassado.

Essa automatização do processo de coleta de dados implica na redução de erros na inserção das informações, visto que diminui o número de etapas em que os dados são processados até chegarem ao seu destino, ou seja, a base de dados do programa de avaliação genética. Além disso, esse processo também é importante do ponto de vista econômico, já que, quanto mais rápida e simples for a alimentação dos dados, menor será o número de pessoas envolvidas nessa tarefa. O tempo

aplicado a este trabalho é igualmente reduzido, possibilitando que os trabalhadores sejam empregados em outras funções de maior produtividade nas fazendas.

Outra etapa de grande relevância em um programa de melhoramento animal é a análise dos dados obtidos. Essa análise ocorre com a utilização de ferramentas complexas que, por muitas vezes, levam períodos consideráveis de tempo (dias) para completar a execução de apenas uma análise. Devido à essa complexidade, é necessário que o usuário possua um conhecimento avançado sobre as ferramentas que realizam comparações e estimativas, pois ele precisa fornecer os parâmetros de entrada em arquivos separados e ainda realizar alterações no próprio código fonte. Desta maneira, um possível erro humano em um arquivo de parâmetros poderia ser descoberto apenas após dias de execução de uma determinada tarefa, ou pior, passar despercebido. Isto poderia ser um agravante na utilização da ferramenta e acarretar em erros no uso do *software* e dispêndio significativo de tempo.

Análises de dados requerem a utilização dos mais diversos tipos de *softwares*, o que requer um constante aprendizado por parte do melhorista ou pesquisador. Esses *softwares*, por sua vez, possuem requisitos específicos como dependências de outros *softwares*, bibliotecas ou até mesmo um sistema operacional específico. Visando centralizar processamentos e otimizar a forma como dados são analisados, existem plataformas como o *Galaxy Project*.

O *Galaxy Project* é uma plataforma *web*, de código aberto, de mineração de dados, gerenciamento de dados e arquivamento eletrônico que visa tornar a bioinformática acessível a pesquisadores com nenhum conhecimento de programação de computador (GOECKS et al., 2010). Essa plataforma dispõe de diversas ferramentas pré-instaladas que realizam vários tipos de análises de dados. Uma das principais características desta ferramenta é a capacidade de inserção de novos *scripts*, que podem rodar nas mais diversas linguagens de programação. A possibilidade de inserir novas ferramentas faz com que o *Galaxy* possa ser um mediador entre o usuário e as ferramentas executadas. Isso elimina diversas etapas na execução da ferramenta, como a criação de arquivos de parâmetros, por exemplo. Esses parâmetros podem ser construídos automaticamente pelo *Galaxy* a partir de formulários intuitivos e a sequência de comandos para iniciar a análise que anteriormente deveria ser executada via terminal não é mais necessária.

Os *scripts* utilizados nas análises do programa de melhoramento animal em questão geralmente são escritos em linguagem de programação *R* ou *Fortran*. Ambas as linguagens podem ser executadas pelo *Galaxy*, deste modo, pode-se utilizar a plataforma como um mediador entre o usuário e o script que será executado, facilitando a forma como o usuário interage com as ferramentas. Além disso, ela acaba eliminando erros que poderiam ser inseridos pelo usuário e que poderiam comprometer os resultados das análises ou exigir diversas tentativas, até que o script simplesmente começasse a ser executado de forma correta. Existe também um apelo muito grande pelo compartilhamento dos resultados e da forma como eles foram obtidos, por isso o *Galaxy* conta com um sistema de compartilhamento de dados e *workflows*. Esses *workflows* mostram de forma visual e detalhada qual foi a ordem de ferramentas, parâmetros e arquivos de dados utilizados a fim de obter os resultados apresentados.

2 Objetivos

2.1 Objetivos gerais

O presente projeto foi executado com o objetivo de desenvolver uma ferramenta para otimizar a forma como são coletados dados e distribuídos os resultados em um programa de melhoramento animal específico. Em uma segunda etapa, foi realizada a importação de um software escrito em linguagem de programação *R* para a plataforma *Galaxy Project*, visando otimizar a forma como a ferramenta é utilizada. Por fim, foi realizada uma análise utilizando o software importado para a plataforma *Galaxy*, com objetivo de verificar as vantagens de utilização e validar a implementação da ferramenta.

2.2 Objetivos específicos

- Desenvolver um aplicativo para dispositivos móveis (sistema operacional iOS) que auxilie na coleta e acesso de dados para o programa de melhoramento animal *PampaPlus*;

- Inserir na plataforma Galaxy o software LD Estimates, com o objetivo de facilitar sua utilização e possibilitando uma série de benefícios que incluem ganho de desempenho, aumento da confiabilidade e compartilhamento de resultados de maneira mais eficaz;
- Executar uma análise com dados reais para validação do software inserido na plataforma Galaxy.

3 PampaPlusApp: Ferramenta para otimizar a coleta de dados utilizados em avaliações genéticas

3.1 Introdução

Em programas de melhoramento genético animal, a coleta de dados se destaca como uma etapa de extrema importância. É preciso cuidado, tanto na coleta, quanto na alimentação do sistema, visto que erros durante o processo de coleta podem comprometer o resultado final da avaliação.

Para realizar essa coleta, os programas de melhoramento possuem técnicos especializados que são credenciados e se fazem presentes nas fazendas. Eles, além de agregarem o conhecimento técnico da avaliação em si, organizam e coordenam o processo de coleta de dados. Esse processo envolve boa parte da equipe das propriedades rurais durante o manejo, pesagens, identificação e verificação dos escores dos animais. Durante o mesmo, são utilizadas planilhas para a tomada de notas, que têm como principais objetivos organizar a coleta e reduzir a entrada de erros. Contudo, cabe ao criador zelar pela fidedignidade dos dados e enviar, geralmente via sistema, ao programa de melhoramento.

Apesar dos criadores poderem utilizar computadores portáteis para realizar o apontamento dos dados, essa não é uma prática comum. Além disso, alguns programas de melhoramento genético animal não possuem ferramentas que trabalhem de forma local (sem conexão com a internet) para o processo de envio de dados. Desta forma, os dados são coletados com auxílio dos técnicos e são enviados posteriormente pelos criadores quando possuem acesso ao sistema informatizado.

Conforme aumenta o número de sub-etapas na coleta, aumentam as chances de erro. Uma simples pesagem pode comprometer uma avaliação, uma

vez que precisa ser lido o valor na balança, anotado em uma planilha e depois inserido no sistema. Essa anotação pode ser ilegível em outro momento, mal compreendida ou até estar fora da faixa de valores aceitáveis. Esse tipo de problema pode ser percebido apenas no momento de transcrição dos dados para o sistema, que pode apresentar uma crítica e impedir o criador de passar um dado errôneo. Contudo, se essa crítica tivesse sido feita no momento da pesagem o erro poderia ter sido esclarecido previamente. Uma possível solução para este gargalo seria providenciar ao criador uma forma de coletar os dados com validação em tempo real diretamente no sistema do programa. Desta forma é eliminado retrabalho e são diminuídas as chances de imputação de erro. Contudo, essa solução ainda precisa lidar com falta de conexão com a internet, restrições ao uso de energia elétrica e proporcionar ao usuário uma experiência de uso que o incentive a utilizar a ferramenta.

Com os avanços recentes da tecnologia no setor de dispositivos móveis que conta com uma imensa variedade de *tablets* e *smartphones*, popularizou-se o uso de aplicativos como solução para os mais diversos tipos de problema. Uma aplicação para dispositivos móveis preenche todos os requisitos necessários para auxiliar o criador tanto na coleta de dados quanto na consulta por informações sobre as avaliações dos seus animais.

O presente projeto foi executado com o objetivo de desenvolver uma ferramenta para otimizar a coleta e acesso a dados fenotípicos de bovinos de corte criados em condições extensivas e avaliados pelo programa de melhoramento animal PampaPlus.

3.2 Escolha da plataforma

O aplicativo para dispositivos móveis foi desenvolvido na plataforma iOS, da Apple. Essa escolha é justificada pela sua alta lucratividade e escalabilidade, uma vez que o sistema operacional da Apple se destaca como uma das principais plataformas de desenvolvimento móvel do mercado. Além disso, possui uma série de recursos que permitem criar aplicativos diferenciados, proporcionando uma ótima experiência ao usuário final (MILANI, 2012).

O *software* foi desenvolvido com a IDE (Ambiente Integrado de Desenvolvimento) gratuita XCODE, disponibilizado pela Apple. Esse ambiente

agiliza o processo de desenvolvimento e garante que o *software* funcione nos mais diversos dispositivos móveis, através de simulações completas e ferramentas de análises de *bugs*. Ademais, a utilização da ferramenta é obrigatória para o *upload* de aplicativos para a AppStore, sendo assim, mesmo que fosse utilizada outra IDE para a construção do aplicativo, a compilação final ainda deveria ser realizada no XCODE. Outra vantagem na utilização do XCODE são recursos como o *auto layout*, que permite que um aplicativo construído para iPhone seja executável também em um iPad, sem grandes esforços. A linguagem de programação utilizada no desenvolvimento do *software* foi *Swift*, uma linguagem de código aberto, desenvolvida pela *Apple* e otimizada para a construção de aplicações para dispositivos móveis.

Esse aplicativo tem um foco especial na experiência final do usuário, utilizando uma interface semelhante a de outros aplicativos altamente difundidos no mercado a fim de facilitar a adaptação do público alvo.

3.3 Desenvolvimento

O primeiro passo para o início do desenvolvimento foi a realização de uma pesquisa e análise dos aplicativos mais conhecidos, com o objetivo de coletar informações sobre os softwares mais populares e porque são aceitos no mercado. Entre esses dados coletados, estão formas de *login*, apresentação e pesquisa de conteúdo em aplicativos, interatividade com a câmera, metodologias para armazenamento de informações no celular e outras.

De posse dos resultados da pesquisa sobre aplicativos de sucesso, foi iniciado o desenvolvimento do aplicativo, realizado em etapas. A primeira etapa visava estabelecer uma forma de conexão entre a base de dados do PampaPlus e o aplicativo em desenvolvimento. A base de dados do PampaPlus é uma base *MySQL*, por isso foi necessário o desenvolvimento de códigos em *PHP* responsáveis pelo *Webservice*. Esse serviço funciona como um intermediário entre o aplicativo e o banco de dados Pampaplus. Ele recebe solicitações via *post* do aplicativo, monta consultas em linguagem *SQL* e retorna ao aplicativo um arquivo no formato *JSON* contendo os dados do banco de dados. Com o acesso à base de dados implementado, o aplicativo começou a tomar forma.

A seguir foi feita uma tela de *login* no aplicativo que valida o acesso de usuários cadastrados no programa PampaPlus por meio de seus respectivos usuários e senhas do site Pampaplusnet. Após ler os dados de login e senha, o aplicativo os envia ao servidor do programa PampaPlus, juntamente com uma chave de sessão criptografada em formato MD5. Sem essa chave, mesmo com o login e senha corretos, o usuário não poderá fazer login. Isso aumenta a segurança na forma em que os dados são trocados entre o aplicativo e o servidor do programa. A Figura 1 mostra visualmente a tela implementada para os usuários efetuarem o *login* no aplicativo. Nesta figura, é possível identificar os campos *login* e senha que devem ser preenchidos para efetuar o login. Em caso de falha, o aplicativo informa ao usuário qual o motivo da falha, que pode ser causado por erro de senha, usuário inválido ou por falta de conexão com a internet.



Figura 1 - Tela de *login* do aplicativo PampaPlusApp. Fonte: Arquivo pessoal.

Juntamente com a construção da tela de *login*, foi também desenvolvida uma tela inicial que age como “*home*”, de onde podem ser iniciados e utilizados quaisquer recursos disponíveis no aplicativo. A Figura 2 apresenta a tela *home* do aplicativo. Nesta tela, pode-se verificar todos os recursos que a aplicação oferece

ao seu usuário. Entre os recursos disponíveis estão “visualizar animais”, “ver e informar dados de desmame” e “sobreano”, além do menu de sincronização de dados entre o aplicativo e o a base de dados do programa de melhoramento animal PampaPlus. Por fim, o menu “ajuda” oferece a visualização de tutoriais de utilização do *software*.

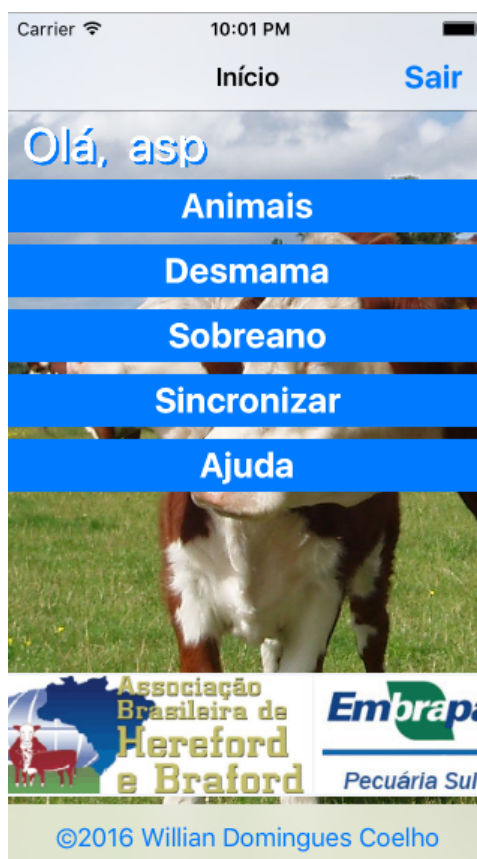


Figura 2 - Tela home do aplicativo PampaPlusApp. Fonte: Arquivo pessoal.

A primeira opção no menu da página *Home* do aplicativo se chama “animais”. Essa opção permite ao usuário visualizar uma lista que contém todos os seus animais registrados no programa de melhoramento genético. Na primeira utilização o aplicativo não terá esses dados, sendo necessário realizar uma sincronização com a base de dados do PampaPlusNet. A Figura 3 exemplifica um primeiro acesso de um usuário à tela de “visualizar animais”. No canto superior direito é possível observar uma opção que leva ao menu de sincronização, se os animais do usuário não estiverem disponíveis podemos utilizar este link para sincronizá-los.

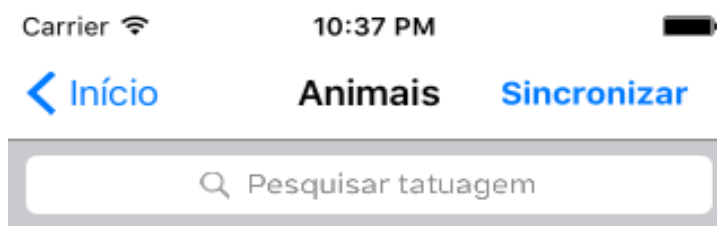


Figura 3 - Consulta de animais no PampaPlusApp. Fonte: Arquivo pessoal.

A Figura 4 exibe o processo de sincronização de animais no aplicativo. É possível observar também as diversas opções de sincronização oferecidas pelo aplicativo, como recuperar dados de desmama e sobreano do servidor para o dispositivo e do dispositivo para o servidor principal. Ainda na Figura 4, pode-se observar o alerta que é lançado no início da sincronização e a barra de acompanhamento da sincronização.



Figura 4 - Tela de sincronização de animais. Fonte: Arquivo pessoal.

Uma vez sincronizadas, todas as informações referentes aos animais deste usuário serão armazenadas no celular, evitando gastos desnecessários de pacotes de dados para carregar esses dados todas as vezes que o usuário precisar. Esse acesso *off-line* é de grande utilidade ao produtor, pois pode consultar informações a qualquer instante, independentemente de sua localização e sem ter que acessar o

sistema online que mantém a base de dados. Podem ser feitas, por exemplo, consultas sobre informações de animais no campo, em ambientes em que não existe conexão com a internet e impróprios para utilização de notebooks. Além disso, existe ainda na ficha de animais um recurso que permite ao usuário coletar fotos dos animais e enviar aos servidores do PpNet, diretamente via app. A Figura 5 exemplifica uma ficha de animal com dados completos de avaliação. Nesta ficha é possível observar também os dados básicos do animal, como nome, tatuagem, sexo, pai, mãe e raça.

Nome: ASP 3014
Gênero: F
Nascimento: 07/02/2003
Pai: Não Consta
Mãe: ASP 8722
Raça: Hereford Mocho

DEPS			
info	DEP	Acurácia	%
PN (Kg)	0.11	13.00	47.00
PDd (Kg)	-2.56	10	64.00
PDM (Kg)	3.91	17	13.00
TMD (Kg)	2.63	-	23.00
PS (Kg)	-1.35	11	58.00
GPD (g/d)	0.60	11	39.00
PMD (Kg)	-7.33	27	24.00
PES (cm)	-0.10		59.00
MSC (1-5)	0.05	12	47.00
EST (1-5)	0.02	12	52.00

PN (Kg)	0.11	13.00	47.00
PDd (Kg)	-2.56	10	64.00
PDM (Kg)	3.91	17	13.00
TMD (Kg)	2.63	-	23.00
PS (Kg)	-1.35	11	58.00
GPD (g/d)	0.60	11	39.00
PMD (Kg)	-7.33	27	24.00
PES (cm)	-0.10		59.00
MSC (1-5)	0.05	12	47.00
EST (1-5)	0.02	12	52.00
CRP (1-5)	0.04	10	43.00
CVD (1-5)	-0.07	22	91.00
UMB (1-5)	0.01	12	42.00
POC (1-5)	-0.01	7.00	51.00

Legenda
 PN = Peso ao Nascer, PDd = Peso ao Desmame, PDM = Peso Desmame Materno, TMD = Total Materno, PS = Peso ao Sobreano, GPD = Ganho Pós-Desmama, PMD = Peso da Mãe ao Desmame, PES = Perímetro Escrotal, MSC = Musculatura, EST = Estatura Corporal, CRP = Estrutura Corporal, CVD = Condição Corporal da Vaca ao Desmame, UMB = Tamanho do Umbigo, POC = Pigmentação Ocular.

Figura 5 - Ficha de um animal do programa PampaPlus. Fonte: Arquivo pessoal.

Cabe destacar que as informações de pai e mãe possuem um recurso adicional, além de mostrar o nome do animal, funcionam como um *link*, ou seja, se o usuário tocar na tatuagem do pai ou na tatuagem da mãe, a ficha deste animal é carregada. Na Figura 5, a tatuagem da mãe está destacada em azul, “ASP 8722”, já

a do pai não consta no banco de dados, por isso, em vez da tatuagem, é exibida a frase “Não Consta”.

Já na Figura 6, pode-se observar que no canto superior esquerdo da tela, onde se localizava o *link* “animais”, visível na Figura 5 e utilizado para retornar a página principal, é substituído pela tatuagem da progênie como forma de retorno rápido. Com isso, podemos ir avançando pela árvore genealógica dos animais e consultando seus respectivos pais e mães, com a possibilidade de fazer o caminho reverso. Para animais base, que não possuem informação de pai, mãe ou ambos, o *link* é inativo.



Figura 6 - Ficha da mãe de um animal do programa PampaPlus. Fonte: Arquivo pessoal.

A segunda opção na tela *Home* é o botão “Desmama”, Figura 7, nesta opção o usuário encontra uma lista de todos os animais que estão aptos a coleta de dados de desmama. A lista, logo que o usuário realiza a instalação do aplicativo, é vazia, sendo necessário fazer a sincronização inicial, semelhante a sincronização de animais. Para realizar a sincronização é preciso acessar a tela conforme exemplificado na Figura 4 e tocar em “Baixar dados para o celular”. Uma vez sincronizados, todos os animais que ainda não possuem registro de dados (escores e outras informações) na desmama e estão aptos para coleta serão exibidos conforme exemplificado na Figura 8. Ainda na Figura 8 podemos visualizar o *link* para realizar a sincronização de dados no canto superior direito e o *link* para

retornar a página inicial no canto superior esquerdo. O usuário também pode optar por procurar um animal digitando sua tatuagem na caixa de busca, na parte superior da tela.



Figura 7 - Botão Desmama destacado. Fonte: Arquivo pessoal.

Tat:	Sexo - Origem
E014	Fêmea - PH
E128	Fêmea - PH
E162	Fêmea - PH
E130	Fêmea - BR
F003	Macho - PH
F007	Macho - PH
F004	Macho - PH
F008	Macho - PH
F009	Fêmea - PH
F449	Macho - PH

Figura 8 - Lista de animais aptos a receber dados de desmama. Fonte: Arquivo pessoal.

Assim que encontrado o animal desejado, utilizando o campo de busca, basta tocar no registro que uma nova tela será aberta. Essa nova tela apresenta um formulário a ser preenchido pelo usuário, respectivo a avaliação do animal em questão. Os campos disponíveis são:

1. Grupo Contemporâneo:

- a. Regime Alimentar
- b. Grupo de Manejo

2. Medidas Objetivas:

- a. Peso

- b. Data da coleta de dados

3. Escores Visuais Obrigatórios:

- a. Musculatura
- b. Estatura
- c. Racial
- d. Aprumos
- e. Características Sexuais

4. Escores Visuais Opcionais:

- a. Estrutura Corporal
- b. Tamanho do Umbigo
- c. Pigmentação Ocular
- d. Castrado
- e. Observações

5. Medidas da Mãe:

- a. Data
- b. Peso da Mãe
- c. Condição Corporal da Mãe

A Figura 9 mostra o formulário por completo na página de inserção de dados do aplicativo para avaliações de desmamas.

< Desmamas Animal F449 Salvar		< Desmamas Animal F449 Salvar	
Características Sexuais	escolha	GRUPO CONTEMPORÂNEO	
ESCORES VISUAIS OPCIONAIS		Regime Alimentar	escolha >
Esutura Corporal	escolha	Grupo de Manejo	Informe o Grupo
MEDIDAS OBJETIVAS			
Tamanho do Umbigo	escolha	Peso(kg)	Informe o peso
Pigmentação Ocular	escolha	Data	Dec 5, 2016
Castrado		ESCORES VISUAIS OBRIGATÓRIOS	
Observações	escolha >	Musculatura	escolha
MEDIDAS DA MÃE		Estatura	escolha
Data		Racial	escolha
Peso Mãe(kg)	Informe o peso	Aprumos	escolha
Condição Corporal da Mãe	escolha		

Figura 9 - Formulário de coleta de dados em desmamas. Fonte: Arquivo pessoal.

A construção de formulários para aplicativos iOS não é trivial, podendo levar um período considerável de tempo para ser executada. Além disso, a construção de formulários mais intuitivos e modernos pode ser ainda mais desafiadora. Em contrapartida, existem diversas bibliotecas de código aberto que automatizam o processo, agilizando assim a construção do aplicativo garantindo uma ótima experiência ao usuário final. Para a construção dos formulários deste aplicativo foi utilizada a biblioteca de código aberto *Eureka*. Essa biblioteca é livre para utilização e possui diversas opções de campos de formulários.

Visando facilitar e tornar mais intuitiva a entrada de informações no formulário de dados de desmamas, alguns dos campos oferecem opções para seleção enquanto outros, que devem ser digitados, verificam os dados informados. Esse tipo de coleta de dados reduz consideravelmente os erros de entrada de dados. Na Figura 10 podemos verificar as opções que são mostradas ao usuário ao tocar no campo regime alimentar. Note que o aplicativo abre uma nova tela que contempla

todas as opções possíveis sem que seja preciso rolar a tela, isso facilita a identificação da opção mais adequada. Além disso, como padrão do aplicativo, surge no canto superior da tela um *link* para voltar ao formulário, sem que seja selecionada a opção de regime alimentar. Esse recurso é particularmente útil quando o usuário está consultando a informação sem desejar alterá-la. Uma vez selecionada a opção desejada, o retorno ao formulário é automático.

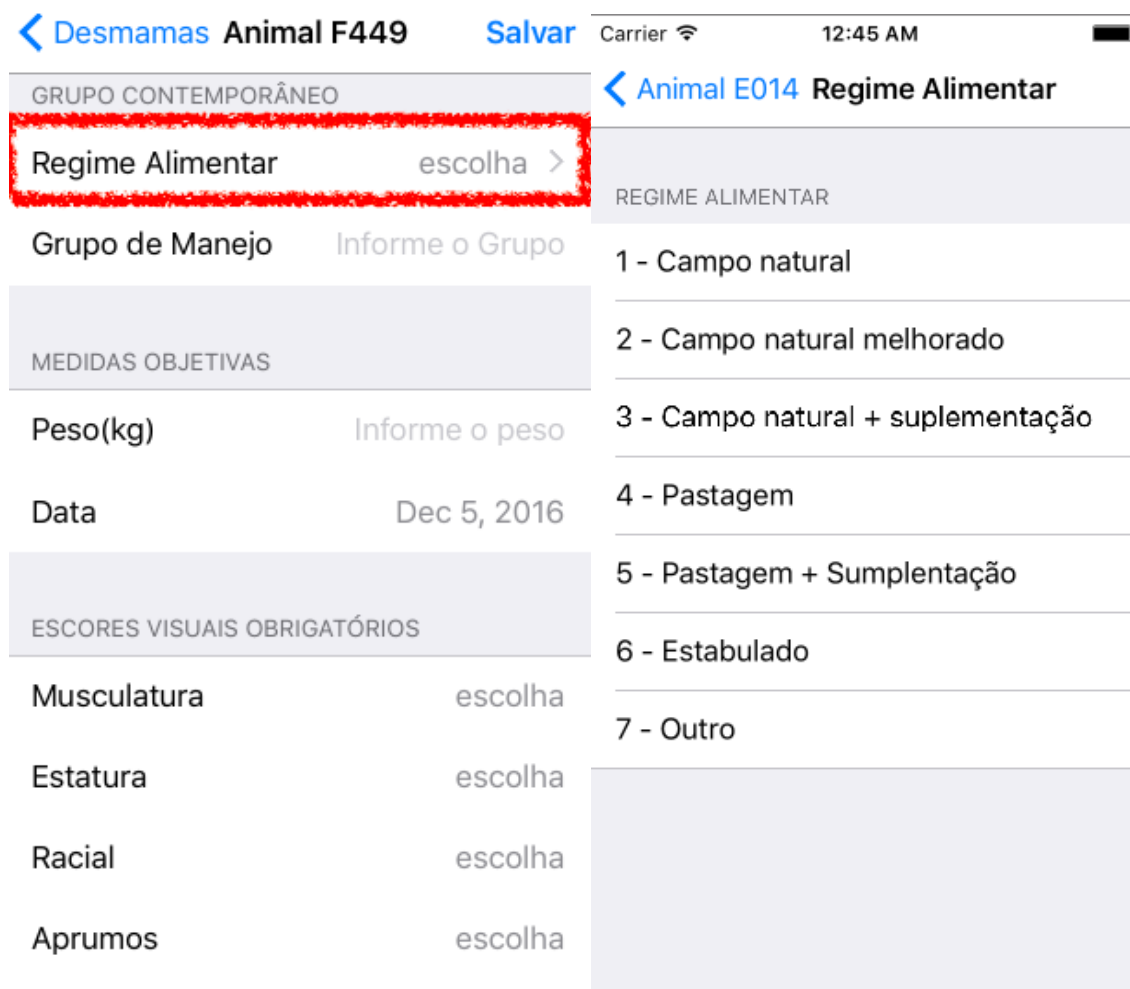


Figura 10 - Tocar na opção regime alimentar, no lado esquerdo, mostra as opções no lado direito da imagem. Fonte: Arquivo pessoal.

Os escores de avaliação dos animais variam entre 1 a 5 e por isso possuem uma seleção mais rápida que a exemplificada no Regime alimentar. Ao tocar no escore que deseja informar, uma caixa é aberta sobre o formulário contendo as notas disponíveis para aquele Escore, junto com uma descrição rápida sobre a nota. Além disso, é apresentada a opção Cancelar, que não altera uma seleção realizada

anteriormente. Após selecionar a opção desejada, o valor é exibido ao lado do Escore na tela de formulário. A Figura 11 exemplifica a escolha do Escore de musculatura e o antes e depois da seleção no formulário.

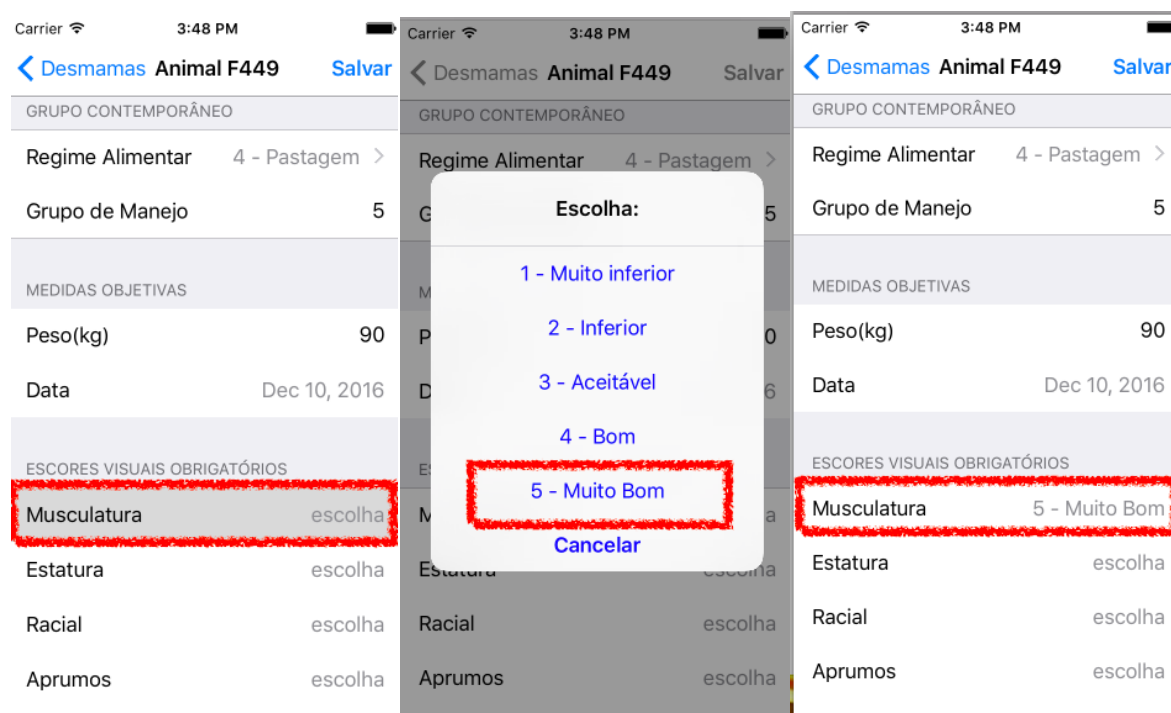


Figura 11 - Informando o Escore de musculatura. Fonte: Arquivo pessoal.

O campo de data da avaliação e a data no “Medidas da Mãe” também possuem uma forma otimizada de coleta, facilitando a seleção da data correta. Na Figura 12 é exemplificada a seleção da Data de avaliação de um animal do programa PampaPlus.

Carrier 9:54 PM

< Desmamas Animal F449 Salvar

Regime Alimentar 1 - Campo nat... >

Grupo de Manejo Informe o Grupo

MEDIDAS OBJETIVAS

Peso(kg) Informe o peso

Data Dec 10, 2016

< > fechar

September	7	2013
October	8	2014
November	9	2015
December	10	2016
January	11	2017
February	12	2018
March	13	2019

Figura 12 - Seleção de data de avaliação de desmama. Fonte: Arquivo pessoal.

Para selecionar a data de avaliação o usuário pode tocar no mês, dia ou ano e girar a “roda” informando rapidamente e de forma intuitiva a data. Uma vez selecionada a data, é preciso tocar em fechar para que o quadro de seleção de data (*Datepicker*) seja oculto. Ainda na Figura 12 é possível observar duas setas “<” e “>” em azul, no lado esquerdo do seletor. Essas setas permitem que o usuário se dirija ao próximo campo a ser preenchido se tocar em “>” ou retorne ao anterior tocando em “<”. Essas setas agilizam ainda mais o processo visto que o campo de seleção de data é ocultado automaticamente. Além disso, ao tocar em uma das setas e se o próximo campo do formulário necessitar de informações escritas, o teclado já será mostrado.

O último tipo de campo no formulário de coleta de dados de avaliação na desmama precisa ser digitado. Um exemplo deste tipo de campo é o peso do

animal. Esse dado é validado em tempo real conforme o usuário o preenche, de modo que sua cor fica vermelha para um dado inválido, ou preta para um dado “na faixa esperada”. Somente números podem ser informados, eliminando a possibilidade do usuário informar acidentalmente um valor não numérico. Na avaliação de desmama o peso varia entre 60kg até 450kg, por isso qualquer valor informado fora desta faixa fica com a cor vermelha, alertando ao usuário que o dado está fora da faixa aceitável. A Figura 13 exemplifica o preenchimento do dado de peso de um animal do programa PampaPlus com a cor vermelha enquanto o peso está fora da faixa (9 kg), e a cor preta quando é um valor aceitável (90 kg).

MEDIDAS OBJETIVAS	
Peso(kg)	9

MEDIDAS OBJETIVAS	
Peso(kg)	90

Figura 13 - Informando peso do animal em avaliação de desmama. Fonte: Arquivo pessoal.

Uma vez preenchido todos os campos do formulário, o resultado será próximo ao exemplo da Figura 14. Tão logo os dados estiverem prontos o usuário precisa salvar as alterações tocando no botão “Salvar” localizado no canto superior direito da tela de formulário, ao lado da tatuagem do animal. Se essa ação não for tomada e o aplicativo for fechado ou ocorra um toque em “< Desmamas” os dados serão perdidos.

GRUPO CONTEMPORÂNEO	
Regime Alimentar	2 - Campo nat... >
Grupo de Manejo	5
MEDIDAS OBJETIVAS	
Peso(kg)	90
Data	Oct 31, 2016
ESCORES VISUAIS OBRIGATÓRIOS	
Musculatura	5 - Muito Bom
Estatura	3 - Médio
Racial	1 - Fora do padrão
Aprumos	3 - Bom

Figura 14 - Exemplo de formulário de avaliação de desmama completo. Fonte: Arquivo pessoal.

Cabe destacar que existem campos **opcionais** e **obrigatórios** na avaliação. Os campos obrigatórios são:

- Regime Alimentar
- Grupo de Manejo
- Peso
- Data

A Figura 15 exemplifica a Ação “Salvar” que abre uma caixa de confirmação perguntando se o usuário realmente deseja guardar os novos dados ou as alterações realizadas na avaliação de desmama do animal em questão. Caso confirme que os dados devem ser guardados, o aplicativo verifica se todos os

campos obrigatórios estão preenchidos e retorna uma das duas opções exibidas na Figura 16, a primeira informando que os dados foram gravados com sucesso e a segunda informando que um ou mais campos obrigatórios não estão preenchidos.



Figura 15 - Tocar no campo salvar exibe uma caixa de confirmação. Fonte: Arquivo pessoal.

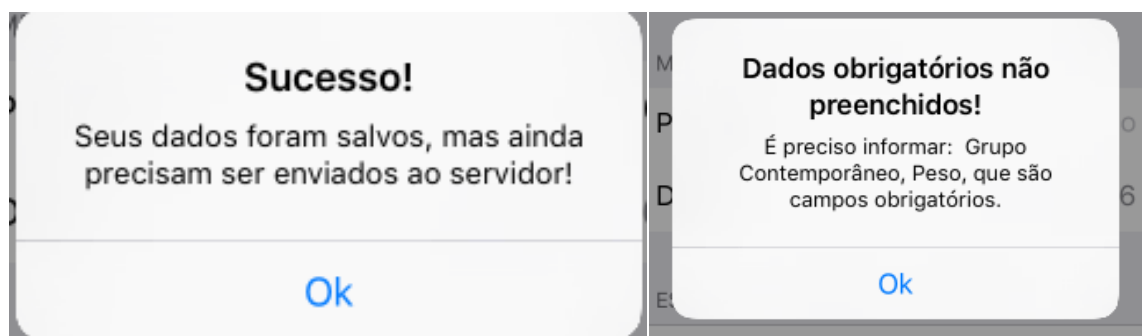


Figura 16 - Retorno do gravar dados validando campos obrigatórios. Fonte: Arquivo pessoal.

Como pode ser verificado na Figura 16, mais especificamente na mensagem de sucesso, os dados estão salvos no dispositivo do usuário, podendo este verificar e consultar as informações de maneira *offline* e a qualquer momento. Contudo, esses dados ainda não estão presentes no servidor do programa, por isso, ainda precisam ser enviados aos servidores do programa.

A sincronização para enviar os dados é semelhante à sincronização para visualizar os animais aptos, ocorrendo na mesma tela do aplicativo. A Figura 17 ilustra o processo de *upload* de dados para os servidores do programa.

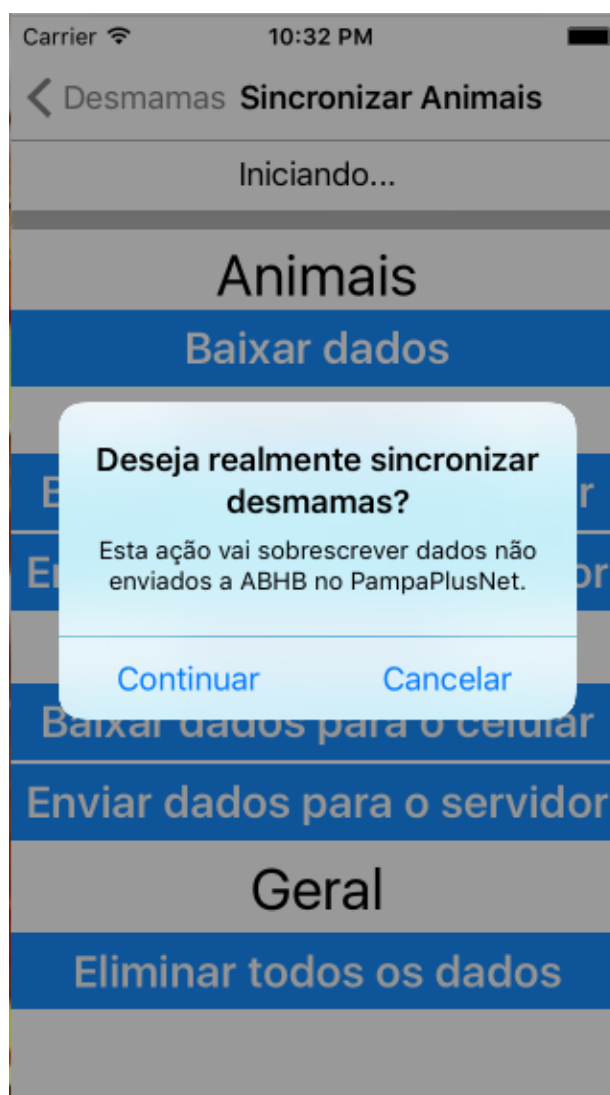


Figura 17 - Enviando dados coletados de Desmama aos servidores PampaPlus. Fonte: Arquivo pessoal.

Com o upload dos dados para os servidores do PampaPlus, se encerra o ciclo de coleta do aplicativo. O próximo passo seria a geração de protocolo resultando na confirmação que os dados foram enviados com sucesso para validação pela ABHB, porém este procedimento está fora do escopo atual do aplicativo e deve ser realizado no site do programa www.pampaplusnet.com.br.

A coleta de dados ao sobreano, outro dos campos do aplicativo, funciona da mesma forma que a da desmama, contando apenas com algumas pequenas diferenças no formulário. Essas diferenças incluem a modificação de alguns campos e a requisição de mais alguns dados de forma obrigatória. Porém, a forma como os animais aptos à avaliação são buscados e o preenchimento dos formulários são idênticos. Como pode ser visualizado na Figura 4, a opção de Sobreanos no menu fica logo após a de Desmamas. Da mesma forma, na Figura 7 observamos que a tela de sincronização de Sobreanos funciona no mesmo local que a de Desmamas.

3.4 Principais Funcionalidades

As funcionalidades do aplicativo são divididas em visualização e alteração de dados. Como exemplificado na seção 4.2, é possível realizar buscas por animais registrados do criador no programa na opção Animais. Essa busca possibilita a visualização de dados, como a data de nascimento, raça, sexo, pai e mãe. Além disso, para os animais que foram avaliados, os resultados da avaliação genética do animal também estarão disponíveis. Já os botões “desmamas” e “sobreanos” resgatam informação e, vão além, permitem que o usuário realize a coleta de dados para avaliação pelo aplicativo, ou seja, é possível enviar informação para o servidor principal do PampaPlusNet. Cabe destacar que estas informações podem ser recolhidas sem que o usuário possua conexão com a internet, podendo então, quando possuir uma conexão, sincronizar os dados ao servidor principal. Esse recurso é o diferencial do aplicativo, que pode facilitar muito a coleta de dados eliminando etapas e evitando a redigitação de dados. Além disso, como ocorrem validações em tempo real, é esperada uma diminuição de erros na coleta de dados.

3.5 Estágio atual

A ferramenta encontra-se plenamente desenvolvida. Os recursos disponíveis atualmente possibilitam para a visualização de informações dos animais e a coleta

de dados para avaliação de desmama e sobreano. Além disso, o aplicativo realiza diversos tipos de validações durante a coleta de dados e apresenta os controles necessários para que as bases do aplicativo e do servidor se mantenham atualizadas e corretas. Todas essas funções foram testadas tanto em simuladores quanto em dispositivos reais. Sendo assim, a ferramenta foi enviada para a loja App Store da Apple, e está disponível para *download* de forma gratuita para qualquer dispositivo que possua o iOS 9 ou superior. Apesar de disponível a qualquer usuário, apenas os usuários do programa poderão utilizar o app, uma vez que o primeiro passo é realizar o *login*.

3.6 Downloads e testes em dispositivos reais

Uma das garantias de controle de qualidade dos aplicativos desenvolvidos para iOS consiste em verificações e testes sobre responsabilidade da própria *Apple*. No momento de submissão de aplicativos para a *App Store* é preciso responder questionários e providenciar um usuário de teste para que o aplicativo possa ser testado pela equipe da *App Store*. Somente após verificado que o aplicativo não possui nenhum tipo de *malware* (*softwares* mal intencionados), e que não apresenta problemas para utilização, o *App* é disponibilizado na loja para *download*. No caso do PampaPlusApp, assim que o aplicativo foi liberado pela *App Store*, ele foi baixado em diversos dispositivos diferentes para realização de testes. Com o objetivo de verificar seu funcionamento, comportamento em dispositivos reais e outras informações cruciais como consumo de banda, processamento e tempo para carregar informações o aplicativo foi baixado em um iPhone 6 de 16GB, um iPhone 5c de 8GB, um iPhone 6s de 64GB, um iPhone 7 de 32GB, um iPad 4 de 32GB e um iPad Air 2 de 128GB.

3.6.1 Armazenamento

O espaçamento de memória ocupado pelo aplicativo em cada dispositivo variou entre 14,5MB (imediatamente após o *download*) e 20MB. Esse acréscimo no espaço de memória ocupado ocorre conforme a quantidade de dados baixados do servidor para o *App* através do recurso de sincronização. Para realizar o *download* do aplicativo é necessário um armazenamento mínimo disponível de 15MB no

dispositivo e conexão com a internet. Além disso, o download da aplicação consumirá 13,5MB ao usuário que efetuar o *download*.

3.6.2 Sincronização de animais, desmamas e sobreanos

Foram utilizados dois usuários de teste do PampaPlus que possuíam registrados 2238 animais e 12975 animais, respectivamente, no programa. Além disso, ambos os usuários possuíam também animais aptos a avaliação de desmama e sobreano, de modo que todos os recursos do aplicativo pudessem ser testados. Foi registrado o tempo médio de *download* dos dados para a sincronização de **todos os animais** para ambos os usuários em duas redes *wi-fi* de velocidades 1MB e 10MB em dois dispositivos distintos um iPhone 6 de 16Gb e iPhone 6s de 64GB. Os resultados desse teste podem ser visualizados na tabela a seguir:

	Wi-fi 1MB – iPhone 6	Wi-fi 10 MB iPhone 6	Wi-fi 1MB iPhone 6s	Wi-fi 10 MB iPhone 6s
Usuário 1 - download de 3 mil animais	58s	62s	53s	50s
Usuário 2 - download de 20 mil animais	656s	651s	349s	345s

Com base na tabela, podemos verificar que a velocidade de conexão não afeta drasticamente o processo de sincronização da base de dados, isso acontece porque na sincronização de dados o processo que consome mais tempo é o armazenamento. Quando o aplicativo dispara uma solicitação ao *webservice*, este devolve um arquivo *json* com os dados dos animais que devem ser salvos. Então a aplicação deve interpretar toda essa informação e começar a gravar de forma a otimizar a consulta na base de dados local. Cabe ressaltar que, quanto maior a capacidade de processamento do dispositivo, mais rápido o processo é concluído. Sendo assim, dispositivos de gerações anteriores podem necessitar de um período de execução maior. Outro ponto importante a ser destacado é que o processo de sincronização de animais só é executado quando uma nova avaliação PampaPlus é liberada, o que ocorre somente a cada dois meses. No que diz respeito a sincronização de dados de desmama e sobreano, não houve diferenças

significativas de tempo para o *download* que foi realizado com sucesso em poucos segundos em todos os ambientes (redes Wi-fi) e dispositivos testados. Isso ocorre porque na segunda sincronização a quantidade de dados é muitas vezes menor que a sincronização de animais, já que o número de animais que estão aptos a avaliação de desmama e sobreano é apenas uma fração do rebanho e a única informação que é buscada são os IDs desses animais.

3.6.3 Consulta de animais e upload de dados

A consulta de animais no aplicativo apresentou os mesmos resultados dos testes em simulador, a tela de animais traz a listagem completa de todo o rebanho do produtor cadastrado no programa PampaPlus em poucos segundos. Ao tocar no menu animais é carregada a tela de animais e o ícone de carregamento é exibido. Foram realizadas consultas em 2 dispositivos distintos com dois usuários teste do PampaPlus com 2238 animais e 12975 animais respectivamente. Neste teste como é acessa a base local de dados a conexão com a internet é dispensável, portanto somente serão diferenciados os tempos de carregamento por dispositivo. O tempo de carregamento necessário para trazer a lista completa de animais é exibido na tabela a seguir:

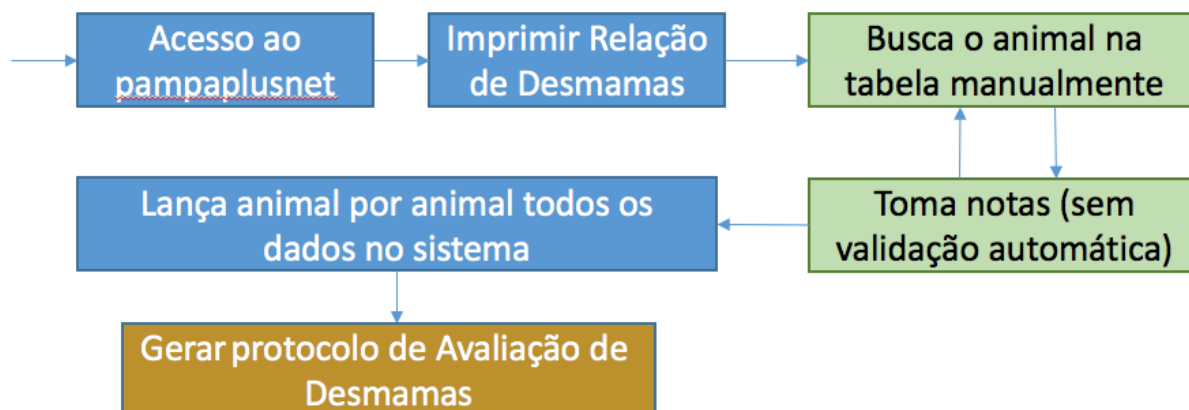
	iPhone 6	iPhone 6s
Usuário 1 - 3 mil animais	2s	1s
Usuário 2 - 12 mil animais	6s	4s

Tão logo a lista é carregada e exibida na tela de animais, foram realizados alguns testes de filtragem por tatuagem e consulta avançada do animal. A filtragem por tatuagem na lista com todos os animais e a consulta da ficha que traz todos os dados do animal apresentaram resultados instantâneos para ambos produtores e dispositivos. A consulta de animais aptos a desmama e ao sobreano tem o mesmo comportamento que a consulta de fichas, porém, devido ao número reduzido de registros, apresentou um tempo de espera muito próximo a 0, por isso não pode ser mensurado.

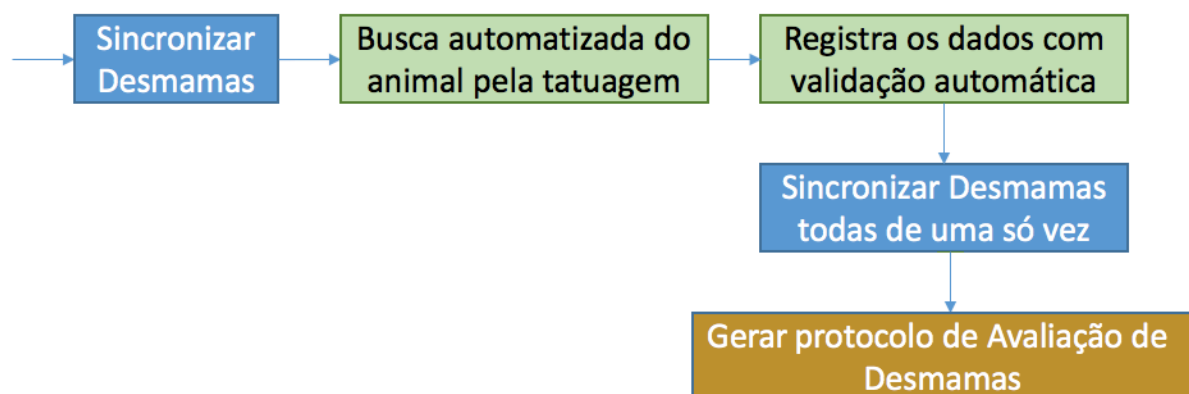
3.7 Comparação entre coleta de dados via PampaPlusApp e convencional

Uma das metas ao proporcionar ao produtor rural uma ferramenta para coleta de dados que caiba no bolso é facilitar a forma como os dados são obtidos. Isso incentiva a permanência no programa e ainda, por meio de validações em tempo real, aumenta a acurácia dos dados coletados. Com o intuito de comparar uma coleta de dados de Desmama do programa PampaPlus com e sem o PampaPlusApp, foram construídos dois fluxogramas. O primeiro considera que os dados serão tomados e registrados em planilhas, impressas ou em um *laptop*, que podem ser exportadas via sistema do programa caracterizando a coleta convencional, enquanto o segundo considera que o produtor possua um *smartphone* ou *tablet* com o PampaPlusApp instalado.

Cada quadro dos fluxogramas representa uma ação que deve ser executada pelo responsável pela coleta dos dados e possuem colorações específicas para destacar a diferença entre os dois processos. No Fluxograma 1, quadros com coloração azul significam acesso/utilização de recursos do sistema pampaplusnet.com.br. Já os quadros com coloração verde significam interação com planilhas impressas ou em um *laptop*. Por fim, o quadro com coloração marrom é a etapa final do processo, quando os dados são confirmados e é gerado um protocolo no sistema pampaplusnet.com.br. No Fluxograma 2, os quadros com coloração azul indicam ações em que o App necessita de conexão com a internet, portanto preferencialmente, devem ser executadas antes da avaliação dos animais em um ambiente com conexão Wi-fi. Os quadros verdes significam ações no aplicativo sem necessidade de conexão com a internet, podendo ser utilizado em qualquer momento, inclusive e preferencialmente durante a avaliação dos animais. Finalmente, o quadro marrom significa geração de protocolo via pampaplusnet.com.br.



Fluxograma 1 – Coleta de dados de Desmama convencional



Fluxograma 2 – Coleta de dados de Desmama utilizando o PampaPlusApp

Comparando os Fluxogramas 1 e 2 é possível identificar prontamente que o número de ações que precisam ser executadas pelo usuário é reduzido pela utilização aplicativo. Além disso, cabe destacar que uma das possíveis desvantagens do aplicativo, que é um preparo anterior à coleta, caracterizado no Fluxograma 2 como **Sincronizar Desmamas**, também ocorre no processo convencional, que seria o acesso ao sistema pampaplusnet.com.br e *download*/impressão da tabela de animais aptos a desmama. A vantagem da utilização do aplicativo que mais se destaca nos Fluxogramas é a eliminação de um retrabalho que ocorre na coleta de dados convencional. Esse retrabalho acontece quando o usuário transcreve os dados coletados para o sistema informatizado do programa, sendo necessário digitar toda a informação obtida, já que o sistema demanda imputação de dados manual e não aceita envio de planilhas. Com a

utilização do aplicativo os dados são inseridos uma única vez no *smartphone* e então sincronizados com o sistema pampaplusnet onde, posteriormente, o usuário deverá acessar e gerar o protocolo de avaliação de desmamas. Outra vantagem de grande impacto no resultado final é a validação em tempo real que ocorre no aplicativo, alertando o usuário sobre possíveis dados coletados de forma errônea. O sistema informatizado pampaplusnet também alertaria, posteriormente, o usuário sobre esses erros no processo convencional de coleta, porém, o aplicativo possui a vantagem de trazer essa informação em um momento mais oportuno para facilitar a correção. Por fim, existe também a praticidade da busca do animal que está em avaliação, uma vez que digitar a tatuagem do animal e encontra-lo de forma instantânea no celular será sempre mais rápido que encontrar em longas listas impressas em papel mesmo que esta lista esteja ordenada. Não foi construído, para fins de comparação, um Fluxograma do procedimento de coleta de dados de Sobreano, pois este é exatamente o mesmo de Desmamas apenas variando os dados que são coletados. Com a comparação dos processos de coleta com e sem o apoio do PampaPlusApp conclui-se que ficam evidenciadas as diversas vantagens que o aplicativo pode proporcionar aos membros do programa de melhoramento genético animal PampaPlus que optarem pela utilização da ferramenta.

3.8 Trabalhos Futuros

Existem diversos recursos do sistema PampaPlusNet que ainda não estão presentes no aplicativo. Alguns desses recursos são consultas de coberturas e descendentes, cadastros de nascimentos, grupos de reprodutores múltiplos e consultas de genealogias. Por isso, a expectativa é que o Aplicativo continue a receber atualizações que acrescentem cada vez mais recursos e o torne uma alternativa completa ao sistema web. Além disso, devido ao grande número de usuários que utilizam o sistema Android em seus dispositivos móveis, pode-se dizer que uma versão para esse sistema se tornará necessária em breve. Porém, esse novo aplicativo já terá uma estrutura pronta para recebê-lo, incluindo todas as tabelas necessárias na base de dados e o *web service*. O que facilitará e tornará o desenvolvimento para a outra plataforma mais ágil.

4 Easy LD-Estimates: adaptação do software *LD-Estimates* a plataforma *Galaxy Project*

4.1 Introdução

Existem diversos *softwares* e *scripts* que auxiliam as análises realizadas por pesquisadores na área de melhoramento genético animal. Contudo, grande parte desses *softwares* foram desenvolvidos com linguagens que não proporcionam recursos que facilitam o desenvolvimento de interface do usuário. Essas linguagens, apesar de apresentarem vantagens para análises estatísticas, precisam ser executadas por linha de comando (via terminal) e sua configuração geralmente requer alteração de arquivos de texto ou até mesmo do código. Devido a isso, a utilização dessas ferramentas requer conhecimentos específicos na área de ciência da computação, o que não é tão comum entre os pesquisadores. Um exemplo desse tipo de software é o *LD Estimate*, cuja otimização de uso será o foco deste trabalho.

O *LD Estimate* foi originalmente desenvolvido na *Michigan State University* em linguagem R por Badke et al (2012). Os scripts, alguns arquivos de dados exemplo, documentação, manual de utilização e outras informações estão disponíveis para download em https://msu.edu/~steibelj/JP_files/LD_estimate.html. Essa ferramenta é de código aberto e é composta por um conjunto de scripts escritos em linguagem R. Ela é utilizada para a estimativa do desequilíbrio de ligação (LD) e persistência de fase. A última versão foi lançada em 2011 e possui uma série de requisitos que restringem o seu uso, cabendo destacar o sistema operacional em que será executada a análise. Atualmente, conforme descrito no site da ferramenta, apenas usuários de Linux ou Mac OS podem executar esses scripts, desde que tenham instalado o interpretador de R e todas as bibliotecas necessárias para a execução. O que, mais uma vez, requer conhecimentos específicos e pode ser um desincentivo ao uso dos scripts. Além disso, considerando que o sistema operacional mais popular do mundo é o Windows nas mais variadas versões, Figura 19, essa limitação deve ser levada em consideração.

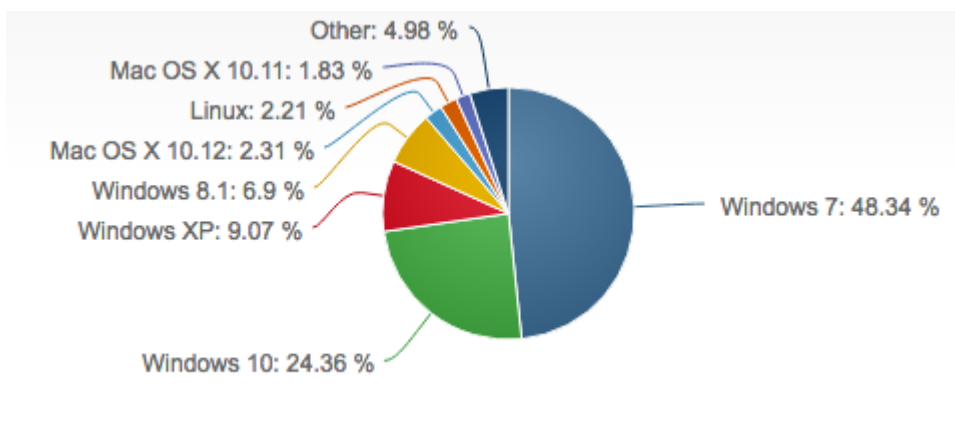


Figura 19 - Desktop Operating System Market Share. Fonte: www.netmarketshare.com.

Outro fator que pode dificultar a difusão da ferramenta é a sua forma de execução. Existe um guia disponível para a utilização da ferramenta em https://msu.edu/~steibelj/JP_files/Ld_user_guide.pdf, onde se pode identificar uma sequência de passos necessários ao usuário para conseguir executar uma análise e então coletar as informações de saída. Visando otimizar a forma como as análises são iniciadas e a forma como os resultados são apresentados ao usuário, o *LD Estimate* foi adaptado a plataforma *Galaxy Project*.

O *Galaxy Project* é uma plataforma web, de código aberto, de mineração de dados, gerenciamento de dados e arquivamento eletrônico que visa tornar a bioinformática acessível a pesquisadores sem conhecimentos de programação de computadores (GOECKS et al., 2010). Essa plataforma dispõe de diversas ferramentas pré-instaladas que realizam vários tipos de análise de dados. Uma das principais características desta ferramenta é a capacidade de inserção de novos scripts, que podem rodar nas mais diversas linguagens de programação. A possibilidade de inserir novas ferramentas faz com que o Galaxy possa ser um mediador entre o usuário e as ferramentas executadas, pois elimina diversas etapas, tais como criação de arquivos de parâmetros, os quais podem ser construídos automaticamente pelo Galaxy a partir de formulários intuitivos. Também evita a sequência de linhas de comando executados via terminal.

Essa adaptação foi realizada com o objetivo de facilitar a utilização dos *scripts* originais do *LD Estimates* de Badke (2012), que necessitam de diversas alterações em código fonte para que o processo de estimação se inicie, podendo dessa forma serem inseridos erros nos códigos fontes, ocasionando retrabalho e demanda de tempo para que o erro seja solucionado. Além disso, uma hierarquia de

diretórios e arquivos de parâmetros precisa ser criada. A execução realizada via terminal pode ser um limitante para usuários não habituados com a interface de linha de comando ou a linguagem estatística R (R Core Team, 2015).

Uma vez instalado na plataforma Galaxy, o *Easy LD Estimates*, nome dado a versão adaptada da ferramenta, constrói a hierarquia de diretórios, cria o arquivo de parâmetros e realiza as alterações necessárias nos scripts de forma totalmente automatizada. Além disso, todos os scripts serão executados na ordem sequencial correta, de modo que as únicas ações tomadas pelo usuário são as de preencher um formulário no início do processo e a de coletar os resultados ao final da análise. Pode-se também integrar outras ferramentas do Galaxy antes ou depois da execução do Easy LD Estimates. Tal integração é feita através do recurso *Workflow* do Galaxy. O Workflow integra as ferramentas enfileirando os processos e sem a necessidade de aguardar um resultado para dar início ao processo seguinte.

O Easy LD Estimates utiliza a estimativa do r^2 como a medida de LD, calculando a média de r^2 para todos os pares de combinações de SNPs, a média de r^2 para os SNPs adjacentes, a correlação de fase dos marcadores em distâncias pré-determinadas, o tamanho efetivo da população e a divergência populacional. Dessa forma a implementação do Easy LD Estimates na plataforma Galaxy permite uma fácil execução e visualização dos dados e resultados por meio de um browser de navegação da web, o que torna o trabalho do pesquisador mais fácil em relação ao uso da interface de linha de comando tradicional.

4.2 Desenvolvimento

Esta seção visa descrever a forma como foi adaptada a ferramenta ao ambiente do *Galaxy Project*, destacando sua utilização, recursos diferenciados do ambiente e as principais diferenças entre a versão original e a versão adaptada. A versão final da ferramenta adaptada está disponível para utilização online no servidor Galaxy do Laboratório Multiusuário de Bioinformática da Embrapa em <https://www.lmb.cnptia.embrapa.br/galaxy/>. Para fins de demonstração, as capturas de telas deste capítulo serão da ferramenta adaptada em funcionamento real no servidor da Embrapa.

Com relação à inclusão de ferramentas, a plataforma Galaxy possui uma estrutura bem definida. Primeiro, é necessário criar um arquivo XML, que irá conter

a interface da ferramenta em questão. Essa ferramenta, na maioria das vezes, se resume a um formulário que define quais serão os arquivos e dados de entrada e de saída. Esse arquivo XML faz a chamada a outros arquivos, que podem ser escritos nas linguagens de programação Python ou Perl, neste segundo arquivo a computação é realizada de fato. Para esta adaptação, o código que constrói o formulário com parâmetros necessários para executar o *Easy LD Estimate* pode ser visualizado no anexo 1.

Na Figura 20, podemos observar a tela de boas vindas do Galaxy Project do LMB da Embrapa, além do menu de ferramentas à esquerda e o histórico de execuções à direita, que será visto nas próximas sessões.



Figura 20 - Página de boas-vindas do Galaxy Project no LMB da Embrapa. Fonte: Arquivo pessoal.

O ambiente Galaxy possui uma série de ferramentas padrão que vêm previamente instaladas. Essas ferramentas são desenvolvidas por uma grande comunidade de pessoas que colaboram com o projeto e são livres para utilização. No ambiente do LMB, existem diversas ferramentas disponíveis, por isso, para encontrar o *Easy LD Estimates*, é preciso acessar o menu *SNV analysis*, onde se encontra a ferramenta, conforme demonstra a Figura 21.

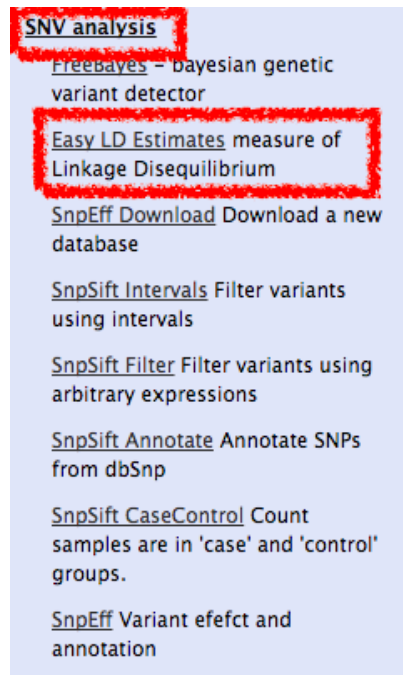


Figura 21 - Local do menu onde se encontra o *Easy LD Estimates*. Fonte: Arquivo pessoal.

Para acessar a ferramenta, basta clicar sobre o *link*, conforme mostra a Figura 21, e então será carregada a página da ferramenta. Essa página, conforme exemplifica a Figura 22, é um formulário de dados de entrada que captura as informações básicas para que a ferramenta possa executar sua análise.

Figura 22 - Visão geral da ferramenta adaptada. Fonte: Arquivo pessoal.

Na Figura 22, pode-se visualizar o formulário de entrada que contém todos os campos necessários para executar o *Easy LD Estimates* no ambiente do *Galaxy Project*. Os campos necessários são:

- Número de cromossomos;
- Arquivo de mapa dos marcadores, que devem ser do tipo SNP (polimorfismos de base única);
- Arquivos de haplótipos (1 por raça);
- Nome de cada raça;
- O vetor de inteiros *Target Distance Id*;
- O vetor de inteiros *Target Distance pp*;
- Marcar verdadeiro ou falso na análise de marcadores esparsos e determinar a densidade dos marcadores dentro do painel caso a opção escolhida for verdadeira.

Na Figura 23, pode-se verificar o campo de inserção de raça expandido, onde deve-se incluir o arquivo e o nome da raça. Para inserir diversas raças basta preencher o nome, selecionar o arquivo de haplótipos e então clicar em *Insert Breed*, que abrirá uma nova caixa (idêntica a anterior) para os novos dados.

The image shows a web interface for adding a breed. It features a header 'Breed' and a list of items. The first item is '1: Breed'. Below this, there is a 'Breed Name' text input field. Underneath is a 'Chromosomes File' section with three buttons: 'Single dataset', 'Multiple datasets', and 'Dataset collection'. A dropdown menu is open, showing '1: Duroc CHR.txt'. At the bottom is a '+ Insert Breed' button.

Figura 23 - Inserção de dados de Raças no *Easy LD Estimates*. Fonte: Arquivo pessoal.

O código XML, responsável pela construção do formulário interativo que serve como interface entre o usuário e a plataforma Galaxy, e que realiza a chamada de execução coletando todos os dados necessários para a execução da

ferramenta de desequilíbrio de ligação está disponível no Anexo 1. Esse código precisa seguir uma estrutura padrão e deve obrigatoriamente:

- Informar qual arquivo realizará o processamento e qual será o interpretador utilizado (*Python* ou *PERL*), linhas 0 e 1 no Anexo 1;
- Conter uma listagem das variáveis (entrada e saída) que serão entregues para processamento, linha 2 até linha a 29 no Anexo 1;
- Deve ser estabelecido quais são as variáveis, vetores, arquivos de entrada, linha 30 até 48 no anexo 1;
- Deve ser estabelecido quais variáveis serão utilizadas para entrega de resultados, informando o nome do arquivo de retorno e o seu formato, linha 49 até 61 no Anexo 1;
- Além disso, neste arquivo, podem ser incluídos exemplos de uso e testes para validação dos dados de entrada e saída, linha 61 até 88 no Anexo 1.

A interface entre usuário e o *LD Estimates* proporcionada pelo ambiente *Galaxy* e implementada via código XML serve como uma forma organizada, unificada e simples tanto para a entrada dos dados quanto para a coleta dos resultados. Porém, é através de um *script* escrito em Perl que a adaptação do *LD Estimates* para a plataforma *Galaxy* é realmente realizada. A ideia por trás do *script* é simples e objetiva: executar o *LD Estimates* **exatamente** como o usuário o executaria, mas de forma automatizada **evitando ao máximo realizar alterações nos *scripts* originais**. A Figura 24 mostra o fluxograma para executar os *scripts* originais, considerando que o usuário esteja utilizando Linux ou Mac OS com interpretador R e todas as bibliotecas dependentes instaladas, destacando todos os processos que devem ser realizados para extrair o máximo de informações.

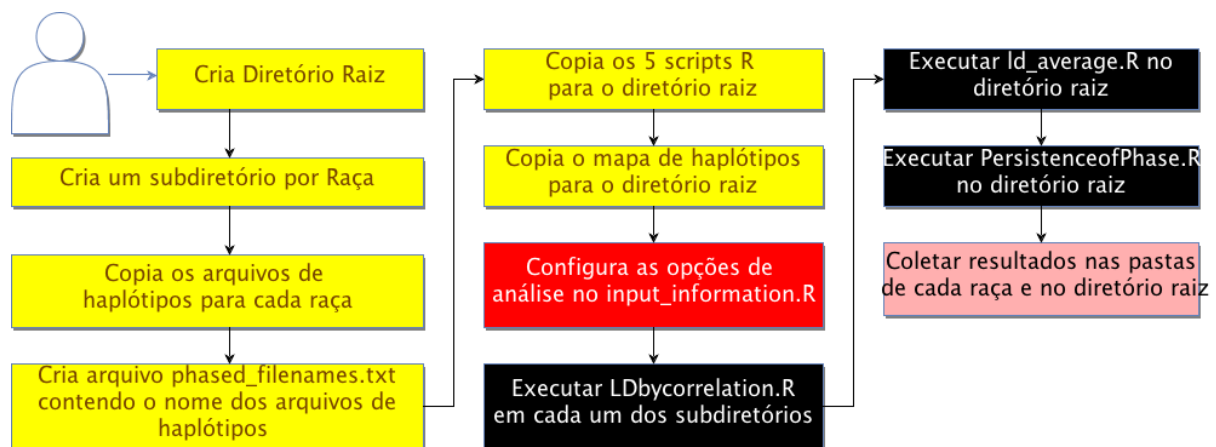


Figura 24 - Fluxograma de uso dos scripts originais. Fonte: Arquivo pessoal.

Ainda na Figura 24, pode-se observar quatro fases distintas durante a utilização da ferramenta, que são destacadas pela coloração dos quadros:

- Quadros coloridos de amarelo fazem parte da construção da hierarquia de diretórios, onde são criados diretórios, subdiretórios, arquivos de texto e copiados os scripts;
- Quadros coloridos de vermelho fazem parte da configuração de execução, onde o código fonte dos *scripts* precisa ser alterado para configurar os parâmetros de entrada da análise;
- Quadros coloridos de preto fazem parte da execução da análise, onde os *scripts* são iniciados por comandos via terminal;
- Quadros coloridos de salmão fazem parte da coleta de resultados, onde os arquivos de saída contendo dados e gráficos devem ser buscados pelo usuário.

A Figura 25 mostra o fluxograma com passos necessários para executar o *Easy LD Estimate* no ambiente *Galaxy*, destacando a redução drástica no número de processos em relação a ferramenta original. Nesta Figura, os quadros brancos representam ações em páginas web. Essa execução via navegador web elimina limitação aos sistemas operacionais Linux e Mac OS, já que quaisquer sistemas operacionais atuais possuem suporte a uma grande variedade de navegadores, como Google Chrome, Mozilla Firefox, Edge, Safari e outros. Essa execução pode

ser realizada inclusive em *tablets*, *smartphones*, *smart TVs* ou qualquer outro dispositivo com um navegador web instalado que suporte os requisitos mínimos do *Galaxy Project*. No quadro salmão, também realizado via navegador, o usuário que iniciou a análise pode visualizar o fazer *download* dos resultados. Esse processo de coleta de resultados também é consideravelmente melhorado na adaptação, trazendo uma série de benefícios ao usuário:

- Visualização rápida dos resultados em qualquer tipo de dispositivo a qualquer momento sem ser necessário *download*, evitando ocupar espaço na memória da máquina que acessa os resultados da consulta;
- Todos os arquivos resultantes da análise disponíveis em único local e não espalhados em diretórios como nos scripts originais;
- Histórico de todas as execuções armazenados em segurança em um servidor protegido;
- Rápido compartilhamento de resultados com outros pesquisadores no próprio ambiente;
- Todos os arquivos temporários para executar a análise são eliminados automaticamente.

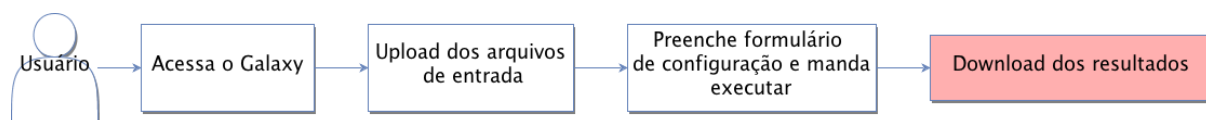


Figura 25 - Fluxograma de execução dos *scripts* no ambiente *Galaxy*. Fonte: Arquivo pessoal.

O *script* em linguagem PERL, que automatiza os procesos e permite a integração do *LD Estimate*, escrito em R, com a plataforma *Galaxy Project* está disponível no Anexo 2. Esse código realiza várias ações que não são visíveis ao usuário utilizando os *scripts* originais, mas automatizando as partes mais críticas do uso do *LD Estimate*. Basicamente, durante sua execução são realizadas as seguintes ações:

- Recebimento dos parâmetros do formulário preenchido pelo usuário;
- Identificação do número total de raças e localização dos arquivos de haplótipos armazenados na base de dados do *Galaxy*;
- Construção da hierarquia de diretórios;

- Construção dos arquivos `input_information.R` em tempo real de acordo com os parâmetros setados pelo usuário no formulário de entrada;
- Chamada de execução do `input_information.R` para cada uma das raças;
- Chamada de execução do `ld_average.R`;
- Chamada de execução do `PersistenceofPhase.R`;
- Atribuição dos arquivos resultantes da análise nas variáveis de retorno ao usuário.

A adaptação dos *scripts* do *LD Estimate* para o ambiente *Galaxy* consiste em automatizar processos manuais sem alterar as características originais da ferramenta. Desta forma, é possível facilitar a utilização da ferramenta enquanto são eliminados passos que poderiam comprometer os resultados da análise por conta de erros do usuário.

4.3 Execução de análise para validação da adaptação

Nesta sessão é realizada uma execução da ferramenta adaptada com o objetivo de validar seu funcionamento. Essa validação é realizada por meio da comparação dos resultados obtidos com a nova ferramenta e os resultados de uma execução do *LD Estimate* original, para um mesmo conjunto de dados de entrada. Além disso, é detalhado todo o processo de execução destacando todos os passos necessários para a obtenção dos resultados. Os dados utilizados são de quatro raças de suínos (Duroc, Hampshire, Landrace e Yorkshire) disponíveis em https://msu.edu/~steibelj/JP_files/LD_estimate.html, no mesmo *site* que são disponibilizados os *scripts* do *LD Estimate*. Em BADKE (2012) é realizada uma análise sobre o mesmo conjunto de dados e são discutidos os resultados, o que está fora do escopo dessa sessão.

O primeiro passo para executar o *Easy LD Estimates* é acessar a ferramenta de upload de dados (*Upload File*), no *Galaxy Project*. Essa ferramenta, geralmente, se encontra no topo do menu de ferramentas, sob o menu *Get Data*, Figura 26. Quando acessada, uma janela do tipo *Pop Up* é exibida ao usuário permitindo que sejam importados arquivos por meio de várias formas distintas como seleção de arquivo local (através de seleção ou arrastando o arquivo em um local específico da página), via transferências (*FTP*) ou colando o conteúdo do arquivo diretamente.

Seja qual for a forma de *upload*, uma vez selecionados os arquivos, é necessário clicar em *start* para que os arquivos sejam transferidos a base do *Galaxy*. A Figura 27 apresenta mostra o *upload* dos arquivos de haplótipos utilizados na execução de validação.

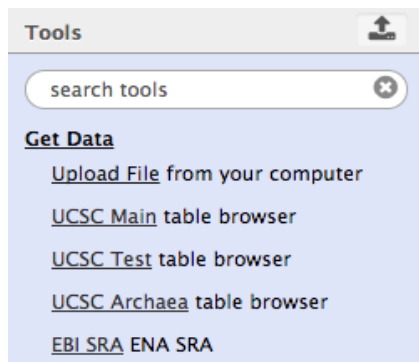


Figura 26 - Ferramentas para *upload* de dados no ambiente *Galaxy*. Fonte: Arquivo pessoal.

Download data directly from web or upload files from your disk

Name	Size	Type	Genome	Settings	Status
Duroc CHR.txt	8.7 MB	Auto-det...	unspecified (?)		100%
Hampshire CHR.txt	7.2 MB	Auto-det...	unspecified (?)		100%
Landrace CHR.txt	9.9 MB	Auto-det...	unspecified (?)		100%
Yorkshire CHR.txt	10.6 MB	Auto-det...	unspecified (?)		58%

Please wait...1 out of 4 remaining.

Choose local file Choose FTP file Paste/Fetch data Start Pause Reset Close

Figura 27 - *Upload* dos arquivos de haplótipos no ambiente *Galaxy*. Fonte: Arquivo pessoal.

Após o *upload* dos arquivos de haplótipos foi realizado o *upload* do arquivo de mapa dos marcadores, necessário para a execução da análise. Tão logo os *uploads* terminam de executar é possível observar, no canto direito da tela, que o histórico de execuções sofreu alterações indicando que os arquivos foram enviados com sucesso e estão disponíveis para utilização. A Figura 28 exibe a condição atual

do histórico de execuções após o *upload* dos arquivos. No histórico de execução estão localizados todos os arquivos que foram enviados ao servidor, permitindo download e visualização do conteúdo.

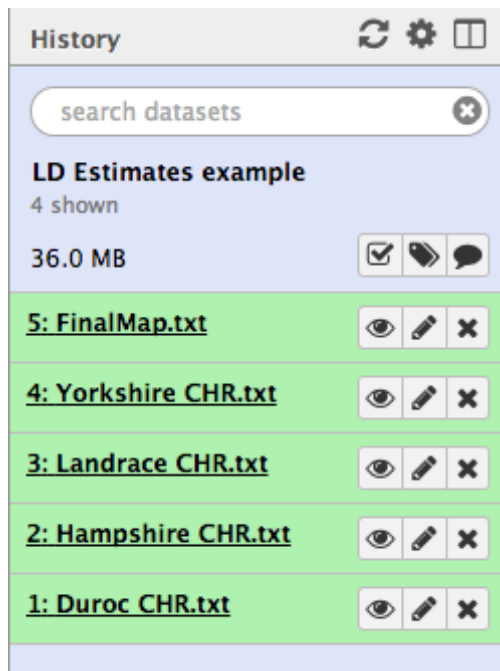


Figura 28 – Histórico de execuções após *upload* dos arquivos de haplótipos e mapa de marcadores.
Fonte: Arquivo pessoal.

O próximo passo é localizar a ferramenta adaptada no menu de ferramentas e iniciar o preenchimento do formulário de parâmetros de entrada. No primeiro campo deve ser selecionado o mapa de marcadores, como este já foi submetido a base de dados está disponível para seleção juntamente com os arquivos de haplótipos. Outros campos foram preenchidos de acordo com os parâmetros utilizados em BADKE (2012). A Figura 29 exhibe parcialmente o formulário preenchido. Além de selecionar o mapa e informar alguns parâmetros numéricos é preciso adicionar raças (*breeds*) e fazer a vinculação com o respectivo arquivo de haplótipos. Cabe ressaltar que o nome dado a cada raça será utilizado nos gráficos e nos arquivos de resultados da análise. A Figura 30 exhibe o vínculo de duas das quatro populações utilizadas para a análise, sendo animais da raça Duroc e Hampshire, respectivamente. Para inserir um número maior de populações basta clicar em *insert breed* para que uma nova caixa de seleção de arquivo será exibida. Para excluir uma das populações basta clicar na lixeira. Na Figura 31 é possível

observar o trecho final do final do formulário e o botão de executar, que deve ser acionado quando o formulário estiver completamente preenchido.

Map file (TXT)

5: FinalMap.txt

Number of chromosomes

18

Target distance Linkage Disequilibrium - example: "45000, 55000, 450000, 550000, 950000, 1050000, 4950000, 5050000"

45000,55000,450000,550000,950000,1050000,4950000,5050000

Target distance Persistence of Phase - example: "0, 10000, 50000, 60000, 100000"

0,10000,50000,60000,100000

Breed

Figura 29 – Primeira parte do formulário de parâmetros preenchido. Fonte: Arquivo pessoal.

Breed

1: Breed

Breed Name

Duroc

Chromossomes File

1: Duroc CHR.txt

2: Breed

Breed Name

Hampshire

Chromossomes File

2: Hampshire CHR.txt

+ Insert Breed

Figura 30 – Vinculação de raças aos arquivos de haplótipos. Fonte: Arquivo pessoal.

Ao clicar em executar os diversos processos que são executados pela ferramenta adaptada são empilhados no servidor para que, assim que possível, sejam executados. Uma tela de confirmação é exibida afirmando que os processos forma adicionados a fila. Além disso, nessa tela de confirmação é dada uma prévia

de todos os arquivos de saída que a análise irá produzir. Na Figura 32 é exibida a tela de confirmação da análise que será executada, onde pode ser conferido ainda todos os arquivos de resultados que são entregues pela ferramenta adaptada ao final da análise.

4: Breed

Breed Name

Yorkshire

Chromossomes File

Single dataset Multiple datasets Dataset collection

4: Yorkshire CHR.txt

+ Insert Breed

Indicate below if you wish to derive sparse set of SNP by consequitively deleting markers from the map

TRUE

Leave it blank if the above option is FALSE, if TRUE - example: "2,4,10"

2,4,10

Execute

Figura 31 – Última parte do formulário de parâmetros. Fonte: Arquivo pessoal.

✓ 1 job has been successfully added to the queue - resulting in the following datasets:

- 1: LD Average_by_dist - PDF
- 2: LD Average_by_chr - PDF
- 3: LD Effective_Pop_Size - PDF
- 4: LD PersistenceofPhase - PDF
- 5: LD average - TXT
- 6: LD sparse - TXT
- 7: LD table1 - TXT
- 8: LD PercentPhasePersistence - TXT
- 9: LD PersistenceofPhase_by_distance - TXT
- 10: LD Timesincepopulationdiverged - TXT
- 11: LD Ne_by_generation - TXT

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

Figura 32 – Tela de confirmação/informações exibida ao executar. Fonte: Arquivo pessoal.

Concomitantemente a exibição da tela de confirmação, novos itens são apresentados no histórico de execuções. Esses itens são os arquivos de resultados da análise em execução. Na Figura 33 é apresentado o histórico de execuções atualizado, onde é possível verificar alguns dos status de execução a partir da coloração da linha. Os status são:

- Cinza – aguardando na fila de execução;
- Amarelo – em execução;
- Verde – concluído;
- Vermelho – erro de execução.

History		↻	⚙	☰
5 shown				
36.0 MB		☑	🗑	💬
🌀	<u>16: LD Ne by generation - TXT</u>	👁	✎	✕
🌀	<u>15: LD Timesincepopulationdiverged - TXT</u>	👁	✎	✕
🌀	<u>14: LD PersistenceofPhase by distance - TXT</u>	👁	✎	✕
🌀	<u>13: LD PercentPhasePersistence - TXT</u>	👁	✎	✕
🌀	<u>12: LD table1 - TXT</u>	👁	✎	✕
🌀	<u>11: LD sparse - TXT</u>	👁	✎	✕
🌀	<u>10: LD average - TXT</u>	👁	✎	✕
🕒	<u>9: LD PersistenceofPhase - PDF</u>	👁	✎	✕
🕒	<u>8: LD Effective Pop Size - PDF</u>	👁	✎	✕
🕒	<u>7: LD Average by chr - PDF</u>	👁	✎	✕
🕒	<u>6: LD Average by dist - PDF</u>	👁	✎	✕

Figura 33 – Histórico de execução após preenchimento do formulário e clique em executar. Fonte: Arquivo pessoal.

Após iniciar a execução, basta acompanhar o histórico de execuções até que todos os itens, relacionados a análise do *Easy LD Estimate*, estejam com coloração verde. O acompanhamento da execução pode ser realizado remotamente, através de qualquer dispositivo que acesse a internet, basta que o usuário faça o seu *login* e poderá verificar o andamento de suas análises. Outro recurso interessante é que em caso de erro na análise, seja por parâmetros incorretos ou inconsistência de dados, o processo pode ser reiniciado a partir de qualquer dispositivo, uma vez que no próprio histórico é possível reiniciar processos alterando parâmetros e arquivos de entrada.

Ao fim da análise, cujo tempo varia de acordo com a carga de trabalho do servidor, o histórico de execução se encontra conforme mostra a Figura 34. Ainda nessa figura é possível verificar no histórico do centro e da direita da imagem dois resultados expandidos. Quando um resultado é expandido, as opções de pré-visualização, *download*, mais informações e outros ficam disponíveis ao usuário. Para visualizar o arquivo sem fazer *download* basta clicar no botão em formato de olho.

The figure displays three panels of the 'LD Estimates example' interface, showing a list of 16 datasets and their execution status. The central and right panels show expanded views of specific datasets with detailed logs and data tables.

Dataset List (Left Panel):

- 16: LD Ne_by_generation - TXT
- 15: LD Timesincepopulationdiverged - TXT
- 14: LD PersistenceofPhase_by_distance - TXT
- 13: LD PercentPhasePersistence - TXT
- 12: LD table1 - TXT
- 11: LD sparse - TXT
- 10: LD average - TXT
- 9: LD PersistenceofPhase - PDF
- 8: LD Effective_Pop_Size - PDF
- 7: LD Average_by_chr - PDF
- 6: LD Average_by_dist - PDF
- 5: FinalMap.txt
- 4: Yorkshire CHR.txt
- 3: Landrace CHR.txt
- 2: Hampshire CHR.txt
- 1: Duroc CHR.txt

Expanded View (Center Panel):

16: LD Ne_by_generation - TXT

15: LD Timesincepopulationdiverged - TXT

7 lines
format: txt, database: ?

starting Duroc... Done.
starting Hampshire... Done.
starting Landrace... Done.
starting Yorkshire... Done.
Done all breeds!
Creating input file... Done.
starting copying files... Done.
starting calculating ld average... Done.
starting calculating persist

Populations	Time since divergence
Duroc-Hampshire	66
Duroc-Landrace	61
Duroc-Yorkshire	65
Hampshire-Landrace	60
Hampshire-Yorkshire	62

14: LD PersistenceofPhase_by_distance - TXT

13: LD PercentPhasePersistence - TXT

12: LD table1 - TXT

11: LD sparse - TXT

10: LD average - TXT

9: LD PersistenceofPhase - PDF

13.0 KB
format: pdf, database: ?

starting Duroc... Done.
starting Hampshire... Done.
starting Landrace... Done.
starting Yorkshire... Done.
Done all breeds!
Creating input file... Done.
starting copying files... Done.
starting calculating ld average... Done.
starting calculating persist

Image in pdf format

8: LD Effective_Pop_Size - PDF

Figura 34 – Histórico de execução após análise concluída. Fonte: Arquivo pessoal.

Com base nos arquivos de saída é possível iniciar a comparação dos resultados obtidos com os resultados da execução do *LD Estimate* original para verificar se existe alguma diferença. A primeira comparação realizada foi entre os gráficos resultantes da análise. As Figuras 35, 36, 37 e 38 apresentam os 4 gráficos lado a lado com os seus respectivos pares gerados pela execução do *LD Estimate original*. Os gráficos no lado esquerdo foram obtidos pelo *Easy LD Estimate*, enquanto os gráficos no lado direito da imagem a partir da ferramenta original. Com base na análise visual dos resultados, nenhuma diferença foi encontrada nesses resultados.

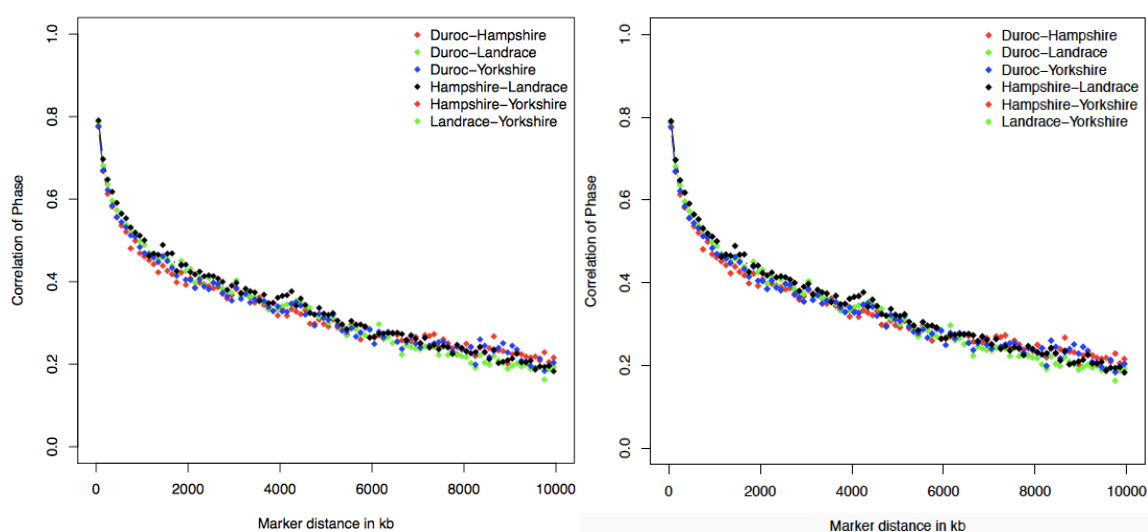


Figura 35 – *PersistenceofPhase*, a esquerda gerado pela execução via *Galaxy* e a direita gerado pelo método de execução original. Fonte: Arquivo pessoal.

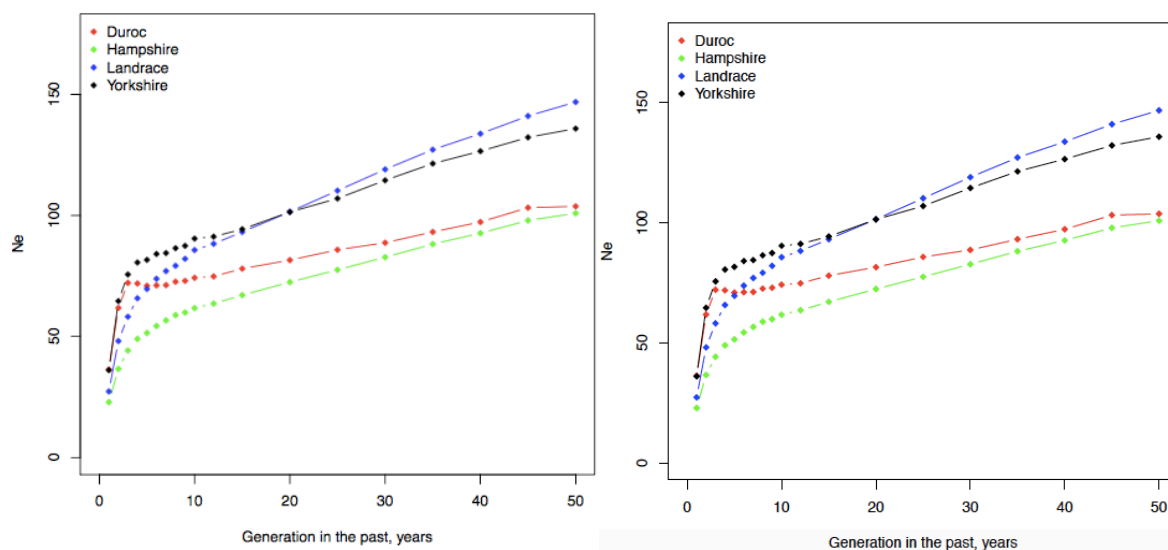


Figura 36 – Gráficos *Effective_Pop_Size*, a esquerda gerado pela execução via *Galaxy* e a direita gerado pelo método de execução original. Fonte: Arquivo pessoal.

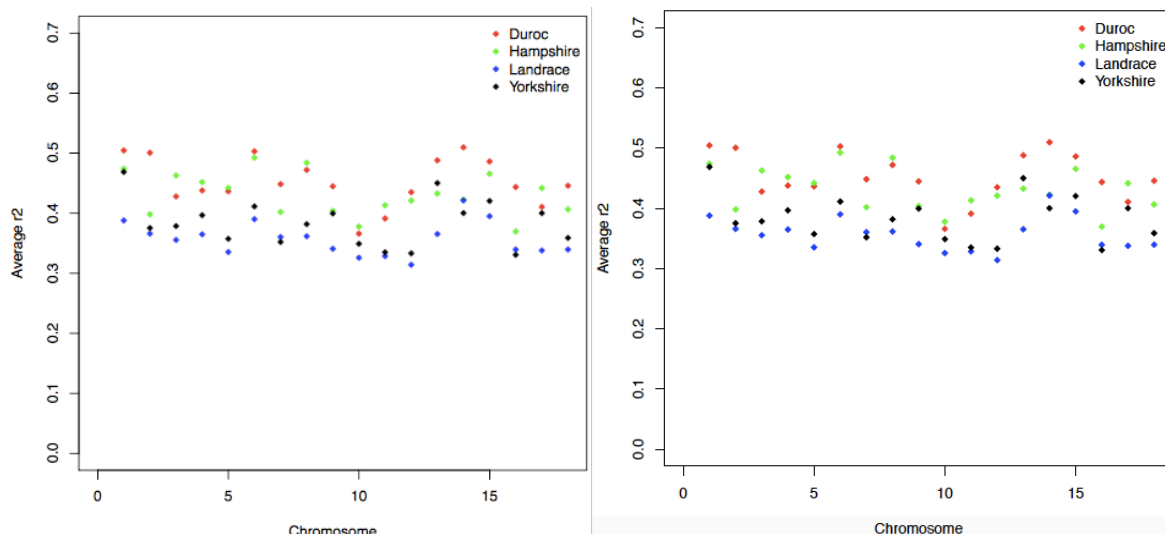


Figura 37 – Gráficos *Average_LD_by_chr*, a esquerda gerado pela execução via *Galaxy* e a direita gerado pelo método de execução original. Fonte: Arquivo pessoal.

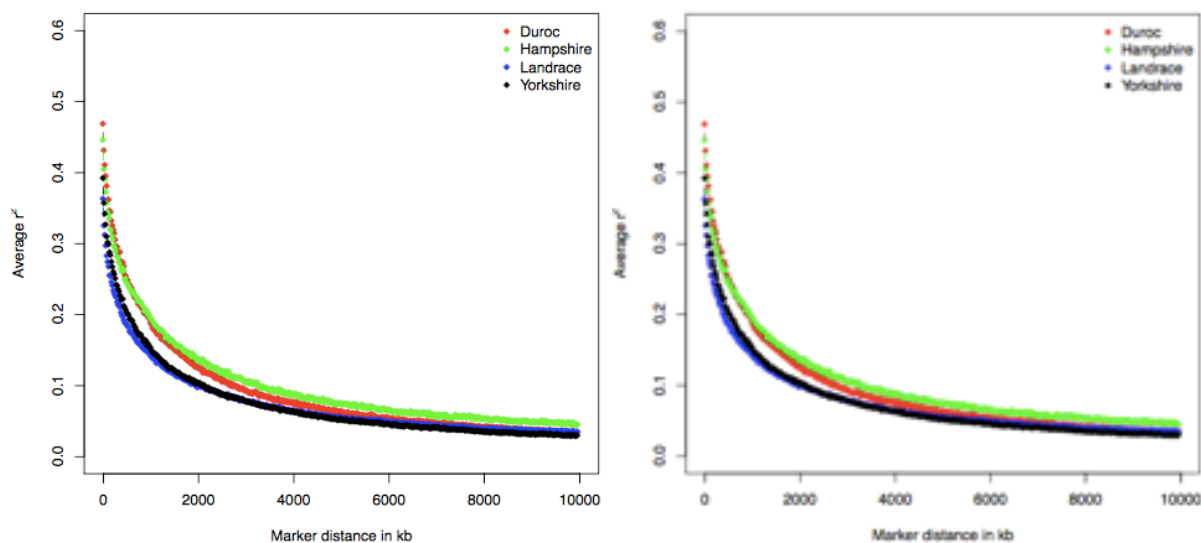


Figura 38 – Gráficos *Average_LD_by_dist*, a esquerda gerado pela execução via *Galaxy* e a direita gerado pelo método de execução original. Fonte: Arquivo pessoal.

A segunda comparação realizada foi verificar se os conteúdos de todos os arquivos de texto gerados possuíam conteúdos idênticos. Assim, foram comparados com utilização do software *Diffchecker* (<https://www.diffchecker.com>) os arquivos: *LD_average*, *LD_sparse*, *LD_table1*, *PercentPhasePersistence*, *PersistenceofPhase_by_distance* e *LD_Ne_by_generation*. Nenhum dos arquivos verificados apresentou diferença entre os que foram gerados na execução via *Galaxy* ou pelo método convencional. Sendo assim, após a comparação dos

resultados considera-se validada a adaptação dos scripts em *R LD Estimate* para a plataforma *Galaxy Project*.

4.4 Benefícios observados

Os principais benefícios que a inclusão da *LD Estimates* na plataforma *Galaxy* proporciona são:

- Execução da ferramenta apenas pelo preenchimento de um simples formulário na plataforma.
- Os arquivos de parâmetros não precisam serem informados manualmente, basta que sejam arrastados para o local certo no *Galaxy Project*, que realizará o upload e os armazenará de forma segura em um servidor na nuvem. Esses arquivos poderão ser utilizados em outras análises futuras sem que o usuário os envie novamente;
- Não requer conhecimentos em *R*, pois não são mais necessárias alterações em códigos fonte;
- A ferramenta é hospedada em um ambiente web, logo pode ser executada a partir de qualquer sistema operacional, inclusive dispositivos móveis como *tablets* e *smartphones*;
- O usuário não precisa de conhecimentos específicos sobre comandos de terminal Linux/Mac OS;
- Não é necessária a instalação das bibliotecas e do *R* no computador do usuário;
- É possível verificar o andamento do processo de forma remota;
- A execução da análise em um servidor de alto rendimento geralmente é mais rápida do que em um PC;
- Os arquivos de saída (resultados) são armazenados no servidor, que é um ambiente seguro e protegido de diversas formas. Aumentando, assim, a garantia de que os dados não serão perdidos;
- Resultados de análises passadas são armazenados e consultados a qualquer momento, evitando gasto desnecessário de memória por parte do usuário;

- Compartilhamento de análises, incluindo workflows. Através dos workflows podem ser observados todos os passos de uma análise, inclusive quais foram os arquivos de entrada;
- Empilhamento de análises, o ambiente do Galaxy Project empilha processos de modo que o pesquisador pode programar que a saída de uma análise seja usada como parâmetro de entrada de outra. Logo, assim que for concluído um processamento inicial, outro processo é disparado e assim sucessivamente até que todos os processos previstos no workflow sejam executados.

5 Considerações Finais

A automatização de etapas em programas de melhoramento genético animal é desafiadora, contudo é possível utilizar as tecnologias disponíveis para combater os gargalos da área. Além disso, a coleta de dados de forma eficiente e com detecção automática de prováveis falhas pode resultar em uma redução drástica de erros, aumentando a acurácia dos resultados. Outro ponto que deve ser considerado é que facilitar o uso das ferramentas de um programa de melhoramento genético animal no lado do produtor pode influenciar outros criadores a aderirem ao programa, enriquecendo ainda mais as avaliações genéticas.

Contudo, não basta facilitar apenas o lado do usuário final dos resultados do programa, a execução de análises requer esforços às vezes desnecessários por parte dos melhoristas e pesquisadores. Devido a isso, a utilização de plataformas de auxílio, como o Galaxy Project, é fundamental. Utilizar servidores para realizar computações que exigem uma demanda maior de análise computacional é de praxe para pesquisadores. Porém, o acompanhamento da análise, que pode levar um período considerável de tempo, requer conhecimentos específicos na área de computação, o que pode ocasionar uma grande frustração ao pesquisador deparando-se com problemas computacionais. A execução em um servidor com Galaxy Project possibilita um acompanhamento remoto simplificado suportado, inclusive, por dispositivos móveis, permitindo ao pesquisador verificar o seu processo sem nenhuma dificuldade. Ademais, o ganho de tempo configurando uma análise também é reduzido consideravelmente.

Em suma, o constante refinamento das etapas na avaliação genética de animais é fundamental para que sejam alcançados resultados ainda mais precisos e maiores ganhos genéticos, com um número maior de pessoas sendo beneficiadas pela produção mais eficiente.

Referências

- ANATEL. Telefonia Móvel – **Dados de acesso em 2015**. Disponível em: <http://www.anatel.gov.br/dados/index.php?option=com_content&view=article&id=270> Acesso em: 29 de junho de 2015.
- BADKE, Y.M.; Bates, R.O.; Ernst, C.W.; Schwab, C. and Steibel, J.P. Estimation of linkage disequilibrium in four US pig breeds. 2012. *BMC Genomics* 2012, **13**:24.
- BLANKENBERG, D; VON KUSTER, G; CORAOR, N; ANANDA, G; LAZARUS, R; MANGAN, M; NEKRUTENKO, A; TAYLOR, J. **Galaxy: a web-based genome analysis tool for experimentalists**. Current Protocols in Molecular Biology. 2010 Jan; Chapter 19:Unit 19.10.1-21.
- GALAXY. Galaxy Project. Disponível em: <<https://galaxyproject.org>> Acesso em: 29 de maio de 2015.
- GIARDINE, B; RIEMER, C; HARDISON, RC; BURHANS, R; ELNITSKI, L; SHAH, P; ZHANG, Y BLANKENBERG, D; ALBERT, I; TAYLOR, J; MILLER, W; KENT, WJ; NEKRUTENKO, A. **Galaxy: a platform for interactive large-scale genome analysis**. Genome Research. 2005 Oct; 15(10):1451-5.
- GOECKS, J; NEKRUTENKO, A; TAYLOR, J; and The Galaxy Team. **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences**. Genome Biol. 2010 Aug 25;11(8):R86.
- IPSOS. Google. **Our Mobile Planet Brazil: understanding the mobile consumer**. Disponível em:< <http://services.google.com/fh/files/misc/omp-2013-br-local.pdf>>May, 2013.
- LECHETA, R. **Desenvolvendo iPhone e iPad**. 3.ed. São Paulo: Novatec Editora, 2014.

MILANI, A. **Programando para iPhone e iPad: aprenda a construir aplicativos para iOS** 1.ed. SP: Ed Novatec, 2012.

SANDVE, G; NEKRUTENKO, A; TAYLOR, J; HOVIG, E. **Ten Simple Rules for Reproducible Computational Research**, in PLoS Computational Biology 9(10): e1003285. doi:10.1371/journal.pcbi.1003285.

SLOGGETT, C; GOONASEKER, N; AFGAN, E. **BioBlend: automating pipeline analyses within Galaxy and CloudMan**, BMC Bioinformatics 2013.

Anexos

Anexo 1 – Código XML para construção do formulário do *Easy LD Estimates*

```

0   <tool id="LD_labegen" name="Easy LD Estimates">
1   <description>measure of Linkage Disequilibrium</description>
2   <command interpreter="perl">
3   LD.pl
4   $map_path
5   $chr
6   $target_distLD
7   $target_distPP
8   $run_sparse
9   $sparse_value
10  $out_file1
11  $out_file2
12  $out_file3
13  $out_file4
14  $out_file5
15  $out_file6
16  $out_file7
17  $out_file8
18  $out_file9
19  $out_file10
20  $out_file11
21  #for $q in $queries
22      ${q.input2}
23  #end for
24  end_breed_names
25  #for $p in $queries
26      ${p.chr_File}
27  #end for
28  end_breed_files
29  </command>

```

```

30  <inputs>
31      <param format="txt" name="map_path" type="data" label="Map file
(TXT)"/>
32      <param name="chr" type="text" label="Number of chromosomes"/>
33      <param name="target_distLD" type="text" size="100" label="Target
distance Linkage Disequilibrium - example: &quot;45000, 55000, 450000, 550000,
950000, 1050000, 4950000, 5050000&quot;" />
34      <param name="target_distPP" type="text" size="100" label="Target
distance Persistence of Phase - example: &quot;0, 10000, 50000, 60000,
100000&quot;"/>
35
36      <repeat name="queries" title="Breed">
37          <param name="input2" type="text" label="Breed Name" />
38          <param format="txt" name="chr_File" type="data"
label="Chromossomes File"/>
39      </repeat>
40
41      <param name="run_sparse" type="select" label="Indicate below if you
wish to derive sparse set of SNP by consecutively deleting markers from the map">
42          <option value="TRUE">TRUE</option>
43          <option value="FALSE">FALSE</option>
44      </param>
45
46      <param name="sparse_value" type="text" size="100" label="Leave it
blank if the above option is FALSE, if TRUE - example: &quot;2,4,10&quot;"/>
47
48  </inputs>
49  <outputs>
50      <data format="pdf" name="out_file1" label="LD Average_by_dist -
PDF"/>
51      <data format="pdf" name="out_file2" label="LD Average_by_chr -
PDF"/>
52      <data format="pdf" name="out_file3" label="LD Effective_Pop_Size -
PDF"/>

```

```

53      <data format="pdf" name="out_file4" label="LD PersistenceofPhase -
PDF"/>
54      <data format="txt" name="out_file5" label="LD average - TXT"/>
55      <data format="txt" name="out_file6" label="LD sparse - TXT"/>
56      <data format="txt" name="out_file7" label="LD table1 - TXT"/>
57      <data          format="txt"          name="out_file8"          label="LD
PercentPhasePersistence - TXT"/>
58      <data          format="txt"          name="out_file9"          label="LD
PersistenceofPhase_by_distance - TXT"/>
59      <data          format="txt"          name="out_file10"         label="LD
Timesincepopulationdiverged - TXT"/>
60      <data format="txt" name="out_file11" label="LD Ne_by_generation -
TXT"/>
61  </outputs>
62  <tests>
63      <test>
64          <output name="out_file1" file="Average_LD_by_dist.pdf"/>
65          <output name="out_file2" file="Average_LD_by_chr.pdf"/>
66          <output name="out_file3" file="Effective_Pop_Size.pdf"/>
67          <output name="out_file4" file="PersistenceofPhase.pdf"/>
68          <output name="out_file5" file="LD_average.txt"/>
69          <output name="out_file6" file="LD_sparse.txt"/>
70          <output name="out_file7" file="LD_table1.txt"/>
71          <output name="out_file8" file="PercentPhasePersistence.txt"/>
72          <output name="out_file9" file="PersistenceofPhase_by_distance.txt"/>
73          <output name="out_file10" file="Timesincepopulationdiverged.txt"/>
74          <output name="out_file11" file="Ne_by_generation.txt"/>
75
76      </test>
77  </tests>
78  <help>
79  .. class:: infomark
80
81  **What it does:**

```

82 The Easy LD Estimates will estimate r^2 as a measure of Linkage Disequilibrium (LD) calculating average r^2 for all pairwise combinations of SNP, average r^2 for adjacent SNP and correlation of phase for markers within a certain distance interval.

83

84 ****example:****

85

86 There is no examples available.

87 </help>

88 </tool>

Anexo 2 - Código PERL para execução do *Easy LD Estimates*

```
0    #!/usr/bin/perl -w
1    use strict;
2    use Cwd 'abs_path';
3    use File::Copy;
4
5    my $map_path = $ARGV[0];
6    my $chr = $ARGV[1];
7    my $distance_id = $ARGV[2];
8    my $distance_pp = $ARGV[3];
9    my $run_sparse = $ARGV[4];
10   my $sparse_value = "";
11   my $counter=5;
12   if($run_sparse ne "FALSE"){
13       $sparse_value = $ARGV[$counter];
14       $counter++;
15   }
16   else{
17       $sparse_value = "2,6,10";
18   }
19   my $out1 = $ARGV[$counter];
20   $counter++;
21   my $out2 = $ARGV[$counter];
22   $counter++;
23   my $out3 = $ARGV[$counter];
24   $counter++;
25   my $out4 = $ARGV[$counter];
26   $counter++;
27   my $out5 = $ARGV[$counter];
28   $counter++;
29   my $out6 = $ARGV[$counter];
30   $counter++;
31   my $out7 = $ARGV[$counter];
```

```

32  $counter++;
33  my $out8 = $ARGV[$counter];
34  $counter++;
35  my $out9 = $ARGV[$counter];
36  $counter++;
37  my $out10 = $ARGV[$counter];
38  $counter++;
39  my $out11 = $ARGV[$counter];
40  $counter++;
41
42  my $breed="\\".$ARGV[$counter]."\\";
43  $counter++;
44  while($ARGV[$counter] ne "end_breed_names"){
45      $breed=$breed.",\\".$ARGV[$counter]."\\";
46      $counter++;
47  }
48  $counter++;
49  my $breed_file="\\".$ARGV[$counter]."\\";
50  $counter++;
51  while($ARGV[$counter] ne "end_breed_files"){
52      $breed_file=$breed_file.",\\".$ARGV[$counter]."\\";
53      $counter++;
54  }
55  my $allbreeds=$breed;
56  my $allbreeds_files=$breed_file;
57  #print $allbreeds;
58  $counter++;
59
60
61
62  #-----Fixing __dq__ problem in
target_dist, sparsing and target
63  # This problem doesn't happen on Mac OS, however when one runs this
script on our galaxy server it happens. No idea why.

```

```

64
65 $distance_ld =~ tr/\__dq__//d; #remove quote
66 $distance_pp =~ tr/\__dq__//d; #remove quote
67 $sparse_value =~ tr/\__dq__//d; #remove quote
68 #-----End fixing __dq__ problem
69
70 #-----breeds
71 my @breed_array = split(/,/,$allbreeds);
72 my $breed_number = @breed_array;
73 my @breed_file_array = split(/,,$allbreeds_files);
74
75 $counter=0;
76 my $LDbycorrelation="";
77
78 foreach my $breed(@breed_array)
79 {
80     $breed =~ tr/"//d; #remove quotes
81     mkdir(abs_path()."/".$breed,0777) || die "can't do mkdir!";
82     $LDbycorrelation=abs_path()."/".$breed."/LDbycorrelation.R";
83
84     print "starting ".$breed."... ";
85
86     #-----creates LDbyCorrelation
87     my $aux ='
88         suppressMessages(library(      ff))
89         map_path="".$map_path."
90         chr<-1:'.$chr.'
91
92     ##Functions used in the code##
93     refer<-function(phased) {
94         b<-sort(unique(phased))[[1]]
95         recoded<-ifelse(phased==b, 1, 0)
96         return(recoded)
97     }

```

```

98
99     filename= '.$breed_file_array[$counter].'
100    pha<-read.table(filename, header=T)
101    map<-read.table(map_path, header=F)
102
103    for (i in chr) {
104        snps_chr <- map[map[,2]==i,1]
105        phased<-as.matrix(pha[pha[,2] %in% snps_chr,-c(1,2)])
106        ph_mat<-t(apply(phased, 1, refer))
107        r<-cor(t(ph_mat))
108        rownames(r)<-pha[pha[,2] %in% snps_chr,2]
109        colnames(r)<-rownames(r)
110        filename<-paste("".abs_path().'.$breed.'/SNP.txt",sep="")
111        write.table(rownames(r),          filename,          row.names=F,
col.names=F, append=T, quote=F)
112        idx<-match(rownames(r),map[,1]) #index for SNP position
in map
113        pos<-map[idx,3]          #snp position
114        for (m in 2:ncol(r)) {          #loop over columns, save upper
triangle of matrix
115            d<-abs(pos[1:(m-1)]-pos[m]) #distance between snp
116            filename<-paste("".abs_path().'.$breed.'/LD.txt", sep="")
117            out_vec<-r[1:(m-1),m]          #output vector: (upper half of)
column m of correlation matrix
118            write.table(cbind(rownames(r)[1:(m-1)],          colnames(r)[m],
out_vec, d , i), filename, col.names=F, row.names=F, quote=F, sep="\t", append=T)
119        }
120        print(paste("DONE WITH CHROMOSOME:", i, sep=" "))
121    }
122    ld_long<-read.table.ffdf(file="".abs_path().'.$breed.'/LD.txt",
colClasses=c("factor","factor","numeric","integer","integer"))
123    ffsave("ld_long",file="".abs_path().'.$breed.'/LD.ff");
124    $counter++;
125    open(my $fh1, '>', $LDbycorrelation) or die "Can't build file $!";

```

```

126     print $fh1 "$aux";
127     close $fh1;
128
129     call_R_func();
130     print("Done.\n")
131 }
132
133     sub call_R_func {
134         my $execute = `Rscript $LDbycorrelation`;
135     }
136     print "Done all breeds!";
137 #-----End Breeds
138
139 #-----Input.R
140 my @colors = ("red","green","blue","black","gold","darkviolet","lightcoral",
"brown","greenyellow","orange4","pink","snow3","tan", "turquoise4", "purple3");
141
142 my $colors_1= "".$colors[0]."";
143
144     for my $i ( 1..$breed_number-1 ) { #create a vector of color depending
on the number of breeds
145         $colors_1 = $colors_1.'.'.$colors[$i]."";
146     }
147     $breed_number = fact($breed_number)/(fact($breed_number-
2)*fact(2));
148     sub fact {
149         my ($n) = @_ ;
150         my $prod = 1;
151         $prod *= $n-- while $n > 0;
152         return $prod;
153     }
154
155     my $colors_2= "".$colors[0]."";
156

```

```
157         for my $i ( 1..$breed_number-1 ) { #create a vector of color depending
on the number os breeds
158             $colors_2 = $colors_2.',".$colors[$i].'';
159         }
160
161         # breed_number!/((breed_number - combinations)! * (combinations)!)
162         print "\nCreating input file... ";
163
164         my $aux ='
165         map_path<-"$.map_path."
166         chr<-1:'.$chr.'
167         breed_fold<-c('.$allbreeds.')
168         breed_name<-c('.$allbreeds.')
169         cols<-c('.$colors_1.')
170         target_dist<-c('.$distance_ld.')
171
172         plot_by_chr<-("Average_LD_by_chr.pdf")
173
174         plot_by_dist<-("Average_LD_by_dist.pdf")
175
176         run_sparse<-'.$run_sparse.'
177
178         sparsing<-c('.$sparse_value.')
179
180         cols2<-c('.$colors_2.')
181         target<-c('.$distance_pp.')
182
183         persist_out<-"PersistenceofPhase_by_distance.txt"
184         phase_out<-"PercentPhasePersistence.txt"
185
186         pdf_out<-"PersistenceofPhase.pdf"
187
188         file_names<-"LD.ff"
189         snp_file<-"SNP.txt"
```

```

190
191     open(my $fh2, '>', "input_information.R") or die "Can't create file $!";
192     print $fh2 "$aux";
193     close $fh2;
194
195     print "Done.";
196
197     #-----
End Input File
198
199     print "\nstarting copying files... ";
200
201     copy ("../../../../tools/Labegen/LDTool/ld_functions.R", abs_path());
202     copy ("../../../../tools/Labegen/LDTool/ld_average.R", abs_path());
203     copy ("../../../../tools/Labegen/LDTool/PersistenceofPhase.R", abs_path());
204     copy ("../../../../tools/Labegen/LDTool/ld_EffectivePopSize.R", abs_path());
205
206     my $PersistenceofPhase = "";
207     $PersistenceofPhase = abs_path()."/PersistenceofPhase.R";
208     #print "\n".$PersistenceofPhase."\n";
209     my $ldAverage = "";
210
211     $ldAverage = abs_path()."/ld_average.R";
212     #print "\n".$ldAverage."\n";
213
214     my $ld_EffectivePopSize = "";
215
216     $ld_EffectivePopSize = abs_path()."/ld_EffectivePopSize.R";
217     #print "\n".$ld_EffectivePopSize."\n";
218
219     #print "\n diretorio corrente=".abs_path()."\n";
220     print "Done.";
221     #-----End Copying
222

```

```
223 print "\nstarting calculating ld average... ";
224
225 call_R_func_ldAverage();
226
227 sub call_R_func_ldAverage {
228     my $execute = `R --no-save -q < $ldAverage`;
229
230 }
231 print "Done.";
232
233 print "\nstarting calculating persistence... ";
234
235 call_R_func_Persistence();
236
237 sub call_R_func_Persistence {
238     my $execute = `R --no-save -q < $PersistenceofPhase`;
239
240 }
241 print "Done.";
242
243 print "\nstarting calculating ld EffectivePopSize... ";
244
245 call_R_func_ld_EffectivePopSize();
246
247 sub call_R_func_ld_EffectivePopSize {
248     my $execute = `R --no-save -q < $ld_EffectivePopSize`;
249 }
250
251 print "Done.";
252 print "\nGenerating out files... ";
253 copy (abs_path()."/Average_LD_by_dist.pdf",$out1);
254 copy (abs_path()."/Average_LD_by_chr.pdf",$out2);
255 copy (abs_path()."/Effective_Pop_Size.pdf",$out3);
256 copy (abs_path()."/PersistenceofPhase.pdf",$out4);
```



```
257 copy (abs_path()"/LD_average.txt",$out5);
258 copy (abs_path()"/LD_sparse.txt",$out6);
259 copy (abs_path()"/LD_table1.txt",$out7);
260 copy (abs_path()"/PercentPhasePersistence.txt",$out8);
261 copy (abs_path()"/PersistenceofPhase_by_distance.txt",$out9);
262 copy (abs_path()"/Timesincepopulationdiverged.txt",$out10);
263 copy (abs_path()"/Ne_by_generation.txt",$out11);
264 print "Done!";
265
266 print "\nDone computing Easy LD Estimates!!\n";
267
```